PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12N 15/16, C07K 14/50, 16/22, A61K 38/22, G01N 33/53, 33/68, C12Q 1/68

A1

(11) International Publication Number: WO 00/60085

(43) International Publication Date: 12 October 2000 (12.10.00)

(21) International Application Number: PCT/US00/08076

(22) International Filing Date: 27 March 2000 (27.03.00)

(30) Priority Data:

60/127,534	2 April 1999 (02.04.99)	US
09/454,470	3 December 1999 (03.12.99)	US

(71) Applicant: MILLENNIUM PHARMACEUTICALS, INC. [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventors: CURTIS, Rory, A., J.; 31 Constitution Drive, Southborough, MA 01772 (US). WRIGHTON, Nicholas, C.; 18 Lloyd Street, Winchester, MA 01890 (US).

(74) Agents: MANDRAGOURAS, Amy, E.; Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109 (US) et al.

(81) Designated States: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: FIBROBLAST GROWTH FACTOR–20

(57) Abstract

The invention provides isolated nucleic acid molecules, designated FGF–20 nucleic acid molecules, which encode novel FGF family members. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing FGF–20 nucleic acid molecules, host cells into which the expression vectors have been introduced, and nonhuman transgenic animals in which an FGF–20 gene has been introduced or disrupted. The invention still further provides isolated FGF–20 proteins, fusion proteins, antigenic peptides and anti–FGF–20 antibodies. Diagnostic methods utilizing compositions of the invention are also provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

FIBROBLAST GROWTH FACTOR–20

Background of the Invention

Growth factors and cytokines regulate a variety of cellular processes including proliferation, differentiation, and morphogenesis during development. Fibroblast growth factor (FGF) was initially characterized as a fibroblast mitogen (Gospodarawicz, D. (1975) *J. Biol. Chem.*, 250:2515-2520). The FGF family currently comprises at least 19 structurally and functionally related proteins, including acidic and basic FGF, FGF-1 and FGF-2 respectively.

Several FGF family members are oncogene products *int2* (FGF-3), *hst* (FGF-4), FGF-5, and *hst2* (FGF-6) (Galzie, Z. *et al.* (1997) *Biochem. Cell. Biol.*, 75:669-685). Other members of this family include keratinocyte growth factor (FGF-7), androgen-induced growth factor (FGF-8) and glia-activating factor (FGF-9) (Galzie, Z. *et al.* (1997) *Biochem. Cell. Biol.*, 75:669-685). FGF-10 is preferentially expressed in the adult lung (Yamasaki, M. *et al.* (1996) *J. Biol. Chem.*, 271:15918-15921). FGFs 11-14, also referred to as FGF homologous factors (FHFs), appear to be involved in the development and function of the nervous system (Smallwood, P.M. *et al.* (1996) *Proc. Natl. Acad. Sci. USA*, 93:9850-9857). FGF-15 displays a regionally restricted and dynamic pattern of expression in the developing nervous system (McWhirter, J.R. *et al.* (1997) *Development*, 124:3221-3232). FGF-16 is predominantly expressed in rat embryonic brown adipose tissue and in the adult heart. FGF-17 displays preferential expression in the neuroepithelia of the isthmus and septum of the embryonic brain (Hoshikawa, M. *et al.* (1998) *Biochem. Biophys. Res. Comm.*, 244:187-191). FGF-18 is expressed primarily in the lungs and kidneys, and stimulates hepatic and intestinal proliferation (Hu, M.C.T. *et al.* (1998) *Mol. Cell. Biol.*, 18:6063-6074). FGF-19 is expressed in the fetal brain (Nishimura, T. *et al.* (1999) *Biochim Biophys Acta*, 1444:148-151).

Target cell responses are mediated, in part, by the binding of FGF ligands to cognate FGF receptors (FGFR) that possess intrinsic tyrosine kinase activity. There are currently four known genes encoding FGF receptors (FGFR-1, FGFR-2, FGFR-3, and FGFR–4), which can give rise to a variety of protein isoforms via alternative RNA splicing (Galzie, Z. *et al.* (1997) *Biochem. Cell. Biol.*, 75:669-685). A given FGFR can

bind different members of the FGF family with varying degrees of specificity. The structure of the FGFR consists of an extracellular region with three immunoglobulin-like domains, a transmembrane region, and a cytosolic tyrosine kinase domain that is activated upon ligand binding. FGF binding causes dimerization of the receptors, resulting in receptor autophosphorylation on tyrosine residues and the activation of intracellular signal transduction cascades. The action of FGF appears to depend on interactions with heparan sulfate proteoglycans in the extracellular matrix. Several proposed roles for proteoglycans in this context include protection from proteolysis, localization, storage, and internalization of growth factors (Faham, S. *et al.* (1998) *Curr. Opin. Struct. Biol.*, 8:578-586). Heparan sulfate proteoglycans may serve as low affinity FGF receptors that act to present FGF to its cognate FGFR, and/or to facilitate receptor oligomerization (Galzie, Z. *et al.* (1997) *Biochem. Cell. Biol.*, 75:669-685).

Summary of the Invention

The present invention is based, at least in part, on the discovery of novel fibroblast growth factor (FGF) family members, referred to herein as "Fibroblast Growth Factor 20" or "FGF-20" nucleic acid and protein molecules. The FGF-20 molecules of the present invention are useful as modulating agents to regulate a variety of cellular processes, including cell proliferation, differentiation, and directed migration. Accordingly, in one aspect, this invention provides isolated nucleic acid molecules encoding FGF-20 proteins or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection of FGF-20-encoding nucleic acids.

In one embodiment, an FGF-20 nucleic acid molecule of the invention is at least 32.2%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, or more identical to the nucleotide sequence (*e.g.*, to the entire length of the nucleotide sequence) shown in SEQ ID NO:1 or 3 or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or a complement thereof.

In one embodiment, an FGF-20 nucleic acid molecule of the invention is at least 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, or more identical to the nucleotide sequence (*e.g.*, to the entire length of the nucleotide

sequence) shown in SEQ ID NO:4, 6, 7 or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or a complement thereof.

In a preferred embodiment, the isolated nucleic acid molecule includes the nucleotide sequence shown SEQ ID NO:1 or 3, or a complement thereof. In another embodiment, the nucleic acid molecule includes SEQ ID NO:3 and nucleotides 533-805 of SEQ ID NO:1. In another preferred embodiment, the nucleic acid molecule consists of the nucleotide sequence shown in SEQ ID NO:1 or 3. In another preferred embodiment, the nucleic acid molecule includes a fragment of at least 107 nucleotides (*e.g.*, 107 contiguous nucleotides) of the nucleotide sequence of SEQ ID NO:1or 3, or a complement thereof.

In another preferred embodiment, the isolated nucleic acid molecule includes the nucleotide sequence shown SEQ ID NO:4 or 6, or a complement thereof. In another embodiment, the nucleic acid molecule includes SEQ ID NO:6 and nucleotides 1-325 of SEQ ID NO:4. In another embodiment, the nucleic acid molecule includes SEQ ID NO:6 and nucleotides 863-2749 of SEQ ID NO:4. In another preferred embodiment, the nucleic acid molecule consists of the nucleotide sequence shown in SEQ ID NO:4 or 6. In another preferred embodiment, the nucleic acid molecule includes a fragment of at least 2329 nucleotides (*e.g.*, 2329 contiguous nucleotides) of the nucleotide sequence of SEQ ID NO:4 or 6, or a complement thereof.

In another preferred embodiment, the isolated nucleic acid molecule includes the nucleotide sequence shown SEQ ID NO:7 or 9, or a complement thereof. In another embodiment, the nucleic acid molecule includes SEQ ID NO:9 and nucleotides 1-1070 of SEQ ID NO:7. In another embodiment, the nucleic acid molecule includes SEQ ID NO:9 and nucleotides 1605-1973 of SEQ ID NO:7. In another preferred embodiment, the nucleic acid molecule consists of the nucleotide sequence shown in SEQ ID NO:7 or 9. In another preferred embodiment, the nucleic acid molecule includes a fragment of at least 1156 nucleotides (*e.g.*, 1156 contiguous nucleotides) of the nucleotide sequence of SEQ ID NO:7 or 9, or a complement thereof.

In another embodiment, an FGF-20 nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence sufficiently identical to the amino acid sequence of SEQ ID NO:2 or an amino acid sequence encoded by the DNA

insert of the plasmid deposited with ATCC as Accession Number _____. In a preferred embodiment, an FGF-20 nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence at least 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more identical to the entire length of the amino acid sequence of SEQ ID NO:2 or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number _____.

In another embodiment, an FGF-20 nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence sufficiently identical to the amino acid sequence of SEQ ID NO:5 or 8, or an amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number _____. In a preferred embodiment, an FGF-20 nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence at least 29.6%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more identical to the entire length of the amino acid sequence of SEQ ID NO:5 or 8, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number _____.

In another preferred embodiment, an isolated nucleic acid molecule encodes the amino acid sequence of monkey or human FGF-20. In yet another preferred embodiment, the nucleic acid molecule includes a nucleotide sequence encoding a protein having the amino acid sequence of SEQ ID NO:2, 5, or 8, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number _____. In yet another preferred embodiment, the nucleic acid molecule is at least 107 nucleotides in length. In a further preferred embodiment, the nucleic acid molecule is at least 107 nucleotides in length and encodes a protein having an FGF-20 activity (as described herein). In yet another preferred embodiment, the nucleic acid molecule is at least 1156 nucleotides in length. In a further preferred embodiment, the nucleic acid molecule is at least 1156 nucleotides in length and encodes a protein having an FGF-20 activity (as described herein).

Another embodiment of the invention features nucleic acid molecules, preferably FGF-20 nucleic acid molecules, which specifically detect FGF-20 nucleic acid molecules relative to nucleic acid molecules encoding non-FGF-20 proteins. For example, in one embodiment, such a nucleic acid molecule is at least 107, 107-150, 150-

200, 200-250, 250-300, 300-350, 350-400, 400-450, 450-500, 500-550, or 550-600, 600-650, 650-700, 700-750, 750-800 or more nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence shown in SEQ ID NO:1, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number_____, or a complement thereof. In another embodiment, such a nucleic acid molecule is at least 1156, 1156-1200, 1200-1400, 1400-1600, 1600-1800, 1800-2000, 2000-2200, 2200-2328, 2329, 2329-2350, 2350-2400, 2400-2450, 2450-2500, 2500-2550, 2550-2600, 2600-2650, 2650-2700 or more nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence shown in SEQ ID NO:4 or 7, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number_____, or a complement thereof.

In preferred embodiments, the nucleic acid molecules are at least 15 (*e.g.*, contiguous) nucleotides in length and hybridize under stringent conditions to nucleotides 1-24, 168-189, 296-301, 552-579, 630-683, or 795-805 of SEQ ID NO:1. In other preferred embodiments, the nucleic acid molecules comprise nucleotides 1-24, 168-189, 296-301, 552-579, 630-683, or 795-805 of SEQ ID NO:1.

In other preferred embodiments, the nucleic acid molecule encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2 or an amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number _____, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID NO:1 or 3 under stringent conditions. In other preferred embodiments, the nucleic acid molecule encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:5 or 8, or an amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number _____, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID NO:4, 6, 7, or 9, under stringent conditions.

Another embodiment of the invention provides an isolated nucleic acid molecule which is antisense to an FGF-20 nucleic acid molecule, *e.g.*, the coding strand of an FGF-20 nucleic acid molecule.

Another aspect of the invention provides a vector comprising an FGF-20 nucleic acid molecule. In certain embodiments, the vector is a recombinant expression vector. In another embodiment, the invention provides a host cell containing a vector of the invention. In yet another embodiment, the invention provides a host cell containing a nucleic acid molecule of the invention. The invention also provides a method for producing a protein, preferably an FGF-20 protein, by culturing in a suitable medium, a host cell, *e.g.*, a mammalian host cell such as a non-human mammalian cell, of the invention containing a recombinant expression vector, such that the protein is produced.

Another aspect of this invention features isolated or recombinant FGF-20 proteins and polypeptides. In one embodiment, the isolated protein, preferably an FGF-20 protein, includes at least one fibroblast growth factor domain. In another embodiment, the isolated protein, preferably an FGF-20 protein, includes a beta trefoil structure. In yet another embodiment, the isolated protein, preferably an FGF-20 protein, includes at least one fibroblast growth factor domain and a beta trefoil structure. In a preferred embodiment, the protein, preferably an FGF-20 protein, includes at least one fibroblast growth factor domain and has an amino acid sequence at least about 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more identical to the amino acid sequence of SEQ ID NO:2, 5, or 8, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number _____. In another preferred embodiment, the protein, preferably an FGF-20 protein, includes a beta trefoil structure and has an amino acid sequence at least about 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more identical to the amino acid sequence of SEQ ID NO:2, 5, or 8, or the amino acid sequence encoded by the DNA insert of the plasmids deposited with ATCC as Accession Numbers _____. In a further preferred embodiment, the protein, preferably an FGF-20 protein, includes at least one fibroblast growth factor domain and a beta trefoil structure and has an amino acid sequence at least about 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more identical to the amino acid sequence of SEQ ID NO:2, 5, or 8, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number _____.

In another preferred embodiment, the protein, preferably an FGF-20 protein, includes at least one fibroblast growth factor domain and plays a role in cell growth, *e.g.*, the regulation of cell proliferation and/or differentiation. In yet another preferred embodiment, the protein, preferably an FGF-20 protein, includes a beta trefoil structure and plays a role in cell growth, *e.g.*, the regulation of cell proliferation and/or differentiation. In a further preferred embodiment, the protein, preferably an FGF-20 protein, includes at least one fibroblast growth factor domain and a beta trefoil structure and plays a role in cell growth, *e.g.*, the regulation of cell proliferation and/or differentiation. In yet another preferred embodiment, the protein, preferably an FGF-20 protein, includes at least one fibroblast growth factor domain and is encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9. In a further embodiment, the protein, preferably an FGF-20 protein, includes a beta trefoil structure and is encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9. In another embodiment, the protein, preferably an FGF-20 protein, includes at least one fibroblast growth factor domain and a beta trefoil structure and is encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 3, 4, 6, 7, or 9.

In another embodiment, the invention features fragments of the protein having the amino acid sequence of SEQ ID NO:2, 5, or 8, wherein the fragment comprises at least 15 amino acids (*e.g.*, contiguous amino acids) of the amino acid sequence of SEQ ID NO:2, 5, or 8, or an amino acid sequence encoded by the DNA insert of the plasmid deposited with the ATCC as Accession Number_____. In another embodiment, the protein, preferably an FGF-20 protein, has the amino acid sequence of SEQ ID NO:2, 5, or 8.

In another embodiment, the invention features an isolated protein, preferably an FGF-20 protein, which is encoded by a nucleic acid molecule consisting of a nucleotide sequence at least about 32.2%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more identical to a nucleotide sequence of SEQ ID NO:1

or 3, or a complement thereof. This invention further features an isolated protein, preferably an FGF-20 protein, which is encoded by a nucleic acid molecule consisting of a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or 3, or a complement thereof.

In another embodiment, the invention features an isolated protein, preferably an FGF-20 protein, which is encoded by a nucleic acid molecule consisting of a nucleotide sequence at least about 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more identical to a nucleotide sequence of SEQ ID NO:4, 6, 7, or 9, or a complement thereof. This invention further features an isolated protein, preferably an FGF-20 protein, which is encoded by a nucleic acid molecule consisting of a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:4, 6, 7, or 9, or a complement thereof.

The proteins of the present invention or portions thereof, *e.g.*, biologically active portions thereof, can be operatively linked to a non-FGF-20 polypeptide (*e.g.*, heterologous amino acid sequences) to form fusion proteins. The invention further features antibodies, such as monoclonal or polyclonal antibodies, that specifically bind proteins of the invention, preferably FGF-20 proteins. In addition, the FGF-20 proteins or biologically active portions thereof can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers.

In another aspect, the present invention provides a method for detecting the presence of an FGF-20 nucleic acid molecule, protein or polypeptide in a biological sample by contacting the biological sample with an agent capable of detecting an FGF-20 nucleic acid molecule, protein or polypeptide such that the presence of an FGF-20 nucleic acid molecule, protein or polypeptide is detected in the biological sample.

In another aspect, the present invention provides a method for detecting the presence of FGF-20 activity in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of FGF-20 activity such that the presence of FGF-20 activity is detected in the biological sample.

In another aspect, the invention provides a method for modulating FGF-20 activity comprising contacting a cell capable of expressing FGF-20 with an agent that modulates FGF-20 activity such that FGF-20 activity in the cell is modulated. In one embodiment, the agent inhibits FGF-20 activity. In another embodiment, the agent stimulates FGF-20 activity. In one embodiment, the agent is an antibody that specifically binds to an FGF-20 protein. In another embodiment, the agent modulates expression of FGF-20 by modulating transcription of an FGF-20 gene or translation of an FGF-20 mRNA. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of an FGF-20 mRNA or an FGF-20 gene.

In one embodiment, the methods of the present invention are used to treat a subject having a disorder characterized by aberrant or unwanted FGF-20 protein or nucleic acid expression or activity by administering an agent which is an FGF-20 modulator to the subject. In one embodiment, the FGF-20 modulator is an FGF-20 protein. In another embodiment the FGF-20 modulator is an FGF-20 nucleic acid molecule. In yet another embodiment, the FGF-20 modulator is a peptide, peptidomimetic, or other small molecule. In a preferred embodiment, the disorder characterized by aberrant or unwanted FGF-20 protein or nucleic acid expression is a disorder associated with deregulated cell growth such as a proliferative or differentiative disorder, including cancer, *e.g.*, carcinoma, sarcoma, or leukemia; tumor angiogenesis and metastasis; skeletal dysplasia; neuronal deficiencies resulting from impaired neural induction and patterning; neurodegenerative disorders, *e.g.*, Alzheimer's disease, dementias related to Alzheimer's disease (such as Pick's disease), Parkinson's and other Lewy diffuse body diseases, multiple sclerosis, amyotrophic lateral sclerosis, progressive supranuclear palsy, epilepsy, Jakob-Creutzfieldt disease, or AIDS related dementia; hepatic disorders; cardiovascular disorders; and hematopoietic and/or myeloproliferative disorders.

The present invention also provides diagnostic assays for identifying the presence or absence of a genetic alteration characterized by at least one of (i) aberrant modification or mutation of a gene encoding an FGF-20 protein; (ii) mis-regulation of the gene; and (iii) aberrant post-translational modification of an FGF-20 protein, wherein a wild-type form of the gene encodes a protein with an FGF-20 activity.

In another aspect the invention provides methods for identifying a compound that binds to or modulates the activity of an FGF-20 protein, by providing an indicator composition comprising an FGF-20 protein having FGF-20 activity, contacting the indicator composition with a test compound, and determining the effect of the test compound on FGF-20 activity in the indicator composition to identify a compound that modulates the activity of an FGF-20 protein.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figure 1 depicts the cDNA sequence and predicted amino acid sequence of monkey FGF-20. The nucleotide sequence corresponds to nucleic acids 1 to 805 of SEQ ID NO:1. The amino acid sequence corresponds to amino acids 1 to 177 of SEQ ID NO: 2. The coding region without the 3' untranslated region of the monkey FGF-20 gene is shown in SEQ ID NO:3.

Figure 2 depicts a structural, hydrophobicity, and antigenicity analysis of the monkey FGF-20 protein.

Figure 3 depicts the results of a search which was performed against the HMM database in which a "Fibroblast growth factor (FGF) domain" was identified in the monkey FGF-20 protein.

Figure 4 depicts a global alignment of the monkey FGF-20 nucleic acid sequence with the Mus musculus mRNA (Accession Number AA175629) using the ALIGN program (version 2.0), using a PAM120 scoring matrix, a gap length penalty of 12 and a gap penalty of 4. The results showed a 32.2% identity between the two sequences.

Figure 5 depicts a global alignment of the monkey FGF-20 protein with the mouse fibroblast growth factor 15 (FGF-15) protein using the ALIGN program (version 2.0), using a PAM120 scoring matrix, a gap length penalty of 12 and a gap penalty of 4. The results showed a 14.7% identity between the two sequences.

Figure 6 depicts a global alignment of the monkey FGF-20 protein with the human fibroblast growth factor 19 (FGF-19) protein using the ALIGN program (version 2.0), using a PAM120 scoring matrix, a gap length penalty of 12 and a gap penalty of 4. The results showed a 17.4% identity between the two sequences.

Figure 7 depicts a local alignment of the monkey FGF-20 protein with the mouse fibroblast growth factor 15 (FGF-15) protein using the LALIGN program (version 2.0u54), using a PAM120 scoring matrix, a gap length penalty of 12 and a gap penalty of 4. The results showed a 35.1% identity between the two sequences over amino acid residues 9-80 of SEQ ID NO:2.

Figure 8 depicts a local alignment of the monkey FGF-20 protein with the human fibroblast growth factor 19 (FGF-19) protein using the LALIGN program (version 2.0u54), using a PAM120 scoring matrix, a gap length penalty of 12 and a gap penalty of 4. The results showed a 39.7% identity between the two sequences over amino acid residues 9-85 of SEQ ID NO:2.

Figure 9 depicts the nucleic acid sequence and predicted amino acid sequence of human FGF-20 as identified within the Homo sapiens 12p13 BAC RPCI11-388F6 genomic fragment (Accession Number AC008012) by homology searching with monkey FGF-20. The nucleotide sequence corresponds to nucleic acids 1 to 2749 of SEQ ID NO:4. The amino acid sequence corresponds to amino acids 1 to 178 of SEQ ID NO: 5. The coding region without the 5' and 3' untranslated regions of the human FGF-20 gene is shown in SEQ ID NO:6.

Figure 10 depicts the cDNA sequence and predicted amino acid sequence of human FGF-20. The nucleotide sequence corresponds to nucleic acids 1 to 1973 of SEQ ID NO:7. The amino acid sequence corresponds to amino acids 1 to 178 of SEQ ID NO: 8. The coding region without the 5' and 3' untranslated regions of the human FGF-20 gene is shown in SEQ ID NO:9.

Figure 11 depicts an alignment of the human FGF-20 cDNA sequence with the human FGF-20 nucleic acid sequence identified within the Homo sapiens 12p13 BAC RPCI11-388F6 genomic fragment (Accession Number AC008012), using the CLUSTAL W (1.74) multiple sequence alignment program.

Figure 12 depicts an alignment of the human FGF-20 protein with the human FGF-20 protein sequence predicted from the Homo sapiens 12p13 BAC RPCI11-388F6 genomic fragment (Accession Number AC008012), using the CLUSTAL W (1.74) multiple sequence alignment program.

Figure 13 depicts a structural, hydrophobicity, and antigenicity analysis of the human FGF-20 protein.

Figure 14 depicts the results of a search which was performed against the HMM database in which a "Fibroblast growth factor (FGF) domain" was identified in the human FGF-20 protein, and the local alignment of the human FGF-20 protein with ProDom entry 549.

Figure 15 depicts a global alignment of the human FGF-20 protein with the human fibroblast growth factor-19 (FGF-19) protein using the GAP program in the GCG software package, using a Blosum 62 matrix and a gap weight of 12 and a length weight of 4. The results showed a 29.6% identity between the two sequences.

Figure 16 depicts a global alignment of the human FGF-20 protein with the mouse fibroblast growth factor-15 (FGF-15) protein using the GAP program in the GCG software package, using a Blosum 62 matrix and a gap weight of 12 and a length weight of 4. The results showed a 22.3% identity between the two sequences.

Figure 17 depicts a global alignment of the human FGF-20 nucleic acid sequence with the monkey FGF-20 nucleic acid sequence using the GAP program in the GCG software package, using a nwsgapdna matrix a gap weight of 12 and a length weight of 4. The results showed a 94.5% identity between the two sequences.

Figure 18 depicts a global alignment of the human FGF-20 protein with the monkey FGF-20 protein using the GAP program in the GCG software package, using a Blosum 62 matrix and a gap weight of 12 and a length weight of 4. The results showed a 93.8% identity between the two sequences.

Detailed Description of the Invention

The present invention is based, at least in part, on the discovery of novel Fibroblast Growth Factor (FGF) family members, referred to herein as "Fibroblast growth factor 20" or "FGF-20" nucleic acid and protein molecules. FGF molecules modulate the proliferation, motility, differentiation, and survival of a variety of cells of

mesodermal, neuroectodermal, ectodermal, and endodermal origin, including fibroblasts, chondrocytes, myoblasts, endothelial cells, astrocytes, neuroblasts, keratinocytes, osteoblasts, and smooth muscle cells (Burgess, W.H. *et al.* (1989) *Ann. Rev. Biochem.*, 58:575-606; Galzie, Z. *et al.* (1997) *Biochem. Cell. Biol.*, 75:669-685). FGF molecules display a broad range of biological activities as mitogens, motogens, angiogenic factors, neurotropic factors, differentiation factors, and oncogenes (Galzie, Z. *et al.* (1997) *Biochem. Cell. Biol.*, 75:669-685). These proteins are important in developmental processes including limb formation, mesoderm induction, and induction and patterning of neural tissues, as well as in the maintenance of tissues and in wound healing and repair.

The FGF-20 molecules of the present invention may also be growth regulatory proteins that function to modulate cell proliferation, differentiation, and motility. Thus, the FGF-20 molecules of the present invention may play a role in cellular growth signaling mechanisms. As used herein, the term "cellular growth signaling mechanisms" includes signal transmission from cell receptors, *e.g.*, growth factor receptors, which regulates 1) cell transversal through the cell cycle, 2) cell differentiation, 3) cell survival, and/or 4) cell migration and patterning. Throughout development and in the adult organism, cell fate and activity is determined, in part, by extracellular and intracellular stimuli, *e.g.*, growth factors, cytokines, hormones, neurotropic factors, angiogenic factors, and chemotactic factors. These stimuli act on their target cells by initiating signal transduction cascades that alter the pattern of gene expression and metabolic activity so as to mediate the appropriate cellular response. The FGF-20 molecules of the present invention may be involved in the initiation of cellular signal transduction pathways that modulate cell growth and differentiation. Thus, the FGF-20 molecules, by participating in cellular growth signaling mechanisms, may modulate cell behavior and act as targets and therapeutic agents for controlling cellular proliferation and differentiation.

Excessive or deficient expression of factors involved in the regulation of signaling pathways associated with cell growth and differentiation can lead to perturbed cellular proliferation, which in turn can lead to cellular proliferative and/or differentiative disorders. As used herein, a "cellular proliferative disorder" includes a disorder, disease, or condition characterized by a deregulated, *e.g.*, upregulated or

downregulated, growth response. As used herein, a "cellular differentiative disorder" includes a disorder, disease, or condition characterized by aberrant or deficient cellular differentiation. Thus, the FGF-20 molecules may act as novel diagnostic targets and therapeutic agents for controlling cellular proliferative and/or differentiative disorders, including cancer, *e.g.*, carcinoma (*e.g.*, colon), sarcoma, leukemia (*e.g.*, erythroleukemia); tumor angiogenesis and metastasis; skeletal dysplasia; hematopoietic and/or myeloproliferative disorders, *e.g.*, anemias (*e.g.*, hemoglobinuria, myelodysplastic syndromes, red cell aplasia, thalassemia), erythrocytosis, neutropenia, neutrophilia, chronic granulomatous disease, eosinophilia, basophilia, monocytosis, histiocytosis, mastocytosis, lymphocytosis, lymphocytopenia, plasmacytosis, thrombocytopenia, thrombocytosis, and lymphoma; hepatic disorders, *e.g.*, cholestasis, cirrhosis, and hyperbilirubinemia; developmental abnormalities associated with aberrant mesodermal patterning; neuronal deficiencies resulting from impaired neural induction and patterning; and neurodegenerative disorders, *e.g.*, Alzheimer's disease, dementias related to Alzheimer's disease (such as Pick's disease), Parkinson's and other Lewy diffuse body diseases, multiple sclerosis, amyotrophic lateral sclerosis, progressive supranuclear palsy, epilepsy, Jakob-Creutzfieldt disease, or AIDS related dementia.

FGF-20-associated or related disorders also include disorders of tissues in which FGF-20 is expressed, *e.g.*, heart, liver, peripheral nervous system (*e.g.*, trigeminal ganglion), and bone marrow.

The FGF-20 molecules of the present invention were identified from a dorsal root ganglion cDNA library. As the dorsal root ganglion contains the cell bodies of sensory neurons involved in pain responses, the FGF-20 molecules of the present invention may also be involved in pain responses. Accordingly, the FGF-20 molecules may also act as novel diagnostic targets and therapeutic agents for controlling pain in a variety of disorders, diseases, or conditions which are characterized by a deregulated, *e.g.*, upregulated or downregulated, pain response. For example, the FGF-20 molecules may provide novel diagnostic targets and therapeutic agents for controlling the exaggerated pain response elicited during various forms of tissue injury, *e.g.*, inflammation, infection, and ischemia, usually referred to as hyperalgesia (described in, for example, Fields, H.L. (1987) *Pain*, New York:McGraw-Hill). Moreover, the FGF-20 molecules may provide novel diagnostic targets and therapeutic agents for

controlling pain, *e.g.*, chronic pain, associated with muscoloskeletal disorders, *e.g.*, joint pain; tooth pain; headaches; neuralgia; pain associated with malignancies, pain associated with surgery, or neuropathic pain.

The FGF-20 molecules of the present invention may also act as novel diagnostic targets and therapeutic agents in cardiovascular disorders such as arteriosclerosis, ischemia reperfusion injury, restenosis, arterial inflammation, vascular wall remodeling, ventricular remodeling, rapid ventricular pacing, coronary microembolism, tachycardia, bradycardia, pressure overload, aortic bending, coronary artery ligation, vascular heart disease, atrial fibrilation, long-QT syndrome, congestive heart failure, sinus node disfunction, angina, heart failure, hypertension, atrial fibrillation, atrial flutter, dilated cardiomyopathy, idiopathic cardiomyopathy, myocardial infarction, coronary artery disease, coronary artery spasm, or arrhythmia.

The term "family" when referring to the protein and nucleic acid molecules of the invention is intended to mean two or more proteins or nucleic acid molecules having a common structural domain or motif and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally or non-naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin, as well as other, distinct proteins of human origin or alternatively, can contain homologues of non-human origin, *e.g.*, monkey proteins. Members of a family may also have common functional characteristics.

For example, sequence conservation among FGF family members indicates that these proteins are likely to include a beta trefoil structure. As used herein, the term "beta trefoil structure" includes a protein tertiary (*i.e.*, three dimensional) structure that preferably has twelve antiparallel beta strands linked to form a structure with three-fold internal symmetry. This structure consists of three copies of a basic four-stranded antiparallel beta sheet. Beta trefoil structures are described in, for example, Zhu, X. *et al.* (1991) *Science*, 251:90-93, the contents of which are incorporated herein by reference.

In another embodiment, an FGF-20 molecule of the present invention is identified based on the presence of a "fibroblast growth factor domain" in the protein or corresponding nucleic acid molecule. As used herein, the term "fibroblast growth factor

domain" includes a protein domain having an amino acid sequence of about 20-100 amino acid residues and having a bit score for the alignment of the sequence to the fibroblast growth factor domain (HMM) of at least 20. Preferably, a fibroblast growth factor domain includes at least about 20-80, or more preferably about 20-60 amino acid residues, and has a bit score for the alignment of the sequence to the fibroblast growth factor domain (HMM) of at least 25, 30, 35, 50 or greater. The fibroblast growth factor domain (HMM) has been assigned the PFAM Accession PF00167 (http://genome.wustl.edu/Pfam/.html). To identify the presence of a fibroblast growth factor domain in an FGF-20 protein, and make the determination that a protein of interest has a particular profile, the amino acid sequence of the protein is searched against a database of HMMs (*e.g.*, the Pfam database, release 2.1) using the default parameters (http://www.sanger.ac.uk/Software/Pfam/HMM_search). A description of the Pfam database can be found in Sonhammer *et al.* (1997) *Proteins* 28(3)405-420 and a detailed description of HMMs can be found, for example, in Gribskov *et al.*(1990) *Meth. Enzymol.* 183:146-159; Gribskov *et al.*(1987) *Proc. Natl. Acad. Sci. USA* 84:4355-4358; Krogh *et al.*(1994) *J. Mol. Biol.* 235:1501-1531; and Stultz *et al.*(1993) *Protein Sci.* 2:305-314, the contents of which are incorporated herein by reference. A search was performed against the HMM database resulting in the identification of a fibroblast growth factor domain in the amino acid sequence of monkey FGF-20 (SEQ ID NO:2) at about residues 1-55 of SEQ ID NO:2. The results of the search are set forth in Figure 3. A fibroblast growth factor domain was also identified in the amino acid sequence of human FGF-20 (SEQ ID NO:5) at about residues 2-56 of SEQ ID NO:5 or 8. The results of the search are set forth in Figure 14.

The fibroblast growth factor domain is characterized by conserved cysteine residues, and in one embodiment comprises the following signature pattern:

G-x-[LI]-x-[STAGP]-x(6,7)-[DE]-C-x-[FLM]-x-E-x(6)-Y (SEQ ID NO:12)

The signature patterns or consensus patterns described herein are described according to the following designation: all amino acids are indicated according to their universal single letter designation; "x" designates any amino acid; x(n) designates n number of amino acids, *e.g.*, x (2) designates any two amino acids, *e.g.*, x (1-3) designates any of one to three amino acids; and, amino acids in brackets indicates any one of the amino

acids within the brackets, *e.g.*, [LI] indicates any of one of either L (leucine) or I (isoleucine). Monkey FGF-20 has such a signature pattern at about amino acids 26 to 50 of SEQ ID NO:2. Human FGF-20 has such a signature pattern at about amino acids 27 to 51 of SEQ ID NO:5 or 8.

The fibroblast growth factor domain comprises a conserved cysteine residue at about amino acid residue 14 of SEQ ID NO:12. Monkey FGF-20 has such a conserved cysteine at about amino acid 39 of SEQ ID NO:2. Human FGF-20 has such a conserved cysteine residue at about amino acid 40 of SEQ ID NO:5 or 8. Alignments of the human FGF-20 protein with the human FGF-19 and mouse FGF-15 proteins (see Figures 15 and 16, respectively), indicate that the conserved cysteine in human FGF-20, at about amino acid 40 of SEQ ID NO:5 or 8, corresponds to cysteine 120 of human FGF-19 and cysteine 127 of mouse FGF-15.

In another preferred embodiment, a fibroblast growth factor domain includes at least about 20-80, or more preferably about 20-60 amino acid residues, and has at least 50-60% homology, preferably about 60-70%, more preferably about 70-80%, or about 80-90% homology with a fibroblast growth factor domain of monkey FGF-20 (residues 1-55 of SEQ ID NO:2) or human FGF-20 (residues 2-56 of SEQ ID NO:5 or 8).

Accordingly, FGF-20 proteins having at least 50-60% homology, preferably about 60-70%, more preferably about 70-80%, or about 80-90% homology with a fibroblast growth factor domain of monkey or human FGF-20 are within the scope of the invention.

Isolated proteins of the present invention, preferably FGF-20 proteins, have an amino acid sequence sufficiently identical to the amino acid sequence of SEQ ID NO:2, 5, or 8 or are encoded by a nucleotide sequence sufficiently identical to SEQ ID NO:1, 3, 4, 6, 7, or 9. As used herein, the term "sufficiently identical" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (*e.g.*, an amino acid residue which has a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences share common structural domains or motifs and/or a common functional activity. For example, amino acid or nucleotide sequences which share common structural domains have at least 30%, 40%, or 50% homology, preferably 60% homology, more preferably 70%-80%, and even more preferably 90-95% homology across

the amino acid sequences of the domains and contain at least one and preferably two structural domains or motifs, are defined herein as sufficiently identical. Furthermore, amino acid or nucleotide sequences which share at least 30%, 40%, or 50%, preferably 60%, more preferably 70-80%, or 90-95% homology and share a common functional activity are defined herein as sufficiently identical.

As used interchangeably herein, an "FGF-20 activity", "biological activity of FGF-20" or "functional activity of FGF-20", refers to an activity exerted by an FGF-20 protein, polypeptide or nucleic acid molecule on an FGF-20 responsive cell or tissue, or on an FGF-20 protein substrate, as determined *in vivo*, or *in vitro*, according to standard techniques. In one embodiment, an FGF-20 activity is a direct activity, such as an association with an FGF-20-target molecule. As used herein, a "target molecule" or "binding partner" is a molecule with which an FGF-20 protein binds or interacts in nature, such that FGF-20-mediated function is achieved. An FGF-20 target molecule can be a non-FGF-20 molecule or an FGF-20 protein or polypeptide of the present invention. In an exemplary embodiment, an FGF-20 target molecule is an FGF-20 substrate, *e.g.*, a FGF receptor or heparan sulfate proteoglycan. Alternatively, an FGF-20 activity is an indirect activity, such as a cellular signaling activity mediated by interaction of the FGF-20 protein with an FGF-20 substrate, *e.g.*, a FGF receptor or heparan sulfate proteoglycan. Preferably, an FGF-20 activity is the ability to act as a growth regulatory factor and to modulate cell proliferation, differentiation, and/or migration.

Accordingly, another embodiment of the invention features isolated FGF-20 proteins and polypeptides having an FGF-20 activity. Preferred proteins are FGF-20 proteins having at least one fibroblast growth factor domain, and, preferably, an FGF-20 activity. Other preferred proteins are FGF-20 proteins having a beta trefoil structure and, preferably, an FGF-20 activity. Yet other preferred proteins are FGF-20 proteins having at least one fibroblast growth factor domain and a beta trefoil structure and, preferably, an FGF-20 activity. Additional preferred proteins have at least one fibroblast growth factor domain and/or a beta trefoil structure, and are, preferably, encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9.

The nucleotide sequence of the isolated monkey FGF-20 cDNA and the predicted amino acid sequence of the monkey FGF-20 polypeptide are shown in Figure 1 and in SEQ ID NOs:1 and 2, respectively. A plasmid containing the nucleotide sequence encoding monkey FGF-20 was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, on _____ and assigned Accession Number _____. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

The partial monkey FGF-20 gene, which is approximately 805 nucleotides in length, encodes a protein having a molecular weight of approximately 20 kD and which is approximately 177 amino acid residues in length.

The nucleotide sequence of the isolated human FGF-20 cDNA and the predicted amino acid sequence of the human FGF-20 polypeptide are shown in Figures 9 and 10 and in SEQ ID NOs:4 and 5, and SEQ ID NOs: 7 and 8, respectively. A plasmid containing the nucleotide sequence encoding human FGF-20 was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, on _____ and assigned Accession Number _____. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

The human FGF-20 gene, which is approximately 1973 nucleotides in length, encodes a protein having a molecular weight of approximately 20 kD and which is approximately 178 amino acid residues in length.

Various aspects of the invention are described in further detail in the following subsections:

I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode FGF-20 proteins or biologically active portions thereof, as well as nucleic acid

fragments sufficient for use as hybridization probes to identify FGF-20-encoding nucleic acid molecules (*e.g.*, FGF-20 mRNA) and fragments for use as PCR primers for the amplification or mutation of FGF-20 nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

The term "isolated nucleic acid molecule" includes nucleic acid molecules which are separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated FGF-20 nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or portion of the nucleic acid sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, as a hybridization probe, FGF-20 nucleic acid molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Moreover, a nucleic acid molecule encompassing all or a portion of SEQ ID NO:1, 3, 4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____ can be isolated by the polymerase chain reaction (PCR) using synthetic oligonucleotide primers designed based upon the sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____.

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to FGF-20 nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:1. The sequence of SEQ ID NO:1 corresponds to the monkey FGF-20 cDNA. This cDNA comprises sequences encoding the monkey FGF-20 protein (*i.e.*, "the coding region", from nucleotides 2-532), as well as 3' untranslated sequences (nucleotides 533-805). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:1 (*e.g.*, nucleotides 2-532, corresponding to SEQ ID NO:3).

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:4. The sequence of SEQ ID NO:4 corresponds to the predicted human FGF-20 cDNA, as identified within the Homo sapiens 12p13 BAC RPCI11-388F6 genomic fragment (Accession Number AC008012) by homology searching with monkey FGF-20. This cDNA comprises sequences encoding the human FGF-20 protein (*i.e.*, "the coding region", from nucleotides 326-862), as well as 5' untranslated sequences (nucleotides 1-325) and 3' untranslated sequences (nucleotides 863-2749). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:4 (*e.g.*, nucleotides 326-862, corresponding to SEQ ID NO:6).

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:7. The sequence of SEQ ID NO:7 corresponds to the human FGF-20 cDNA. This cDNA comprises

sequences encoding the human FGF-20 protein (*i.e.*, "the coding region", from nucleotides 1071-1604), as well as 5' untranslated sequences (nucleotides 1-1070) and 3' untranslated sequences (nucleotides 1605-1973). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:7 (*e.g.*, nucleotides 1071-1604, corresponding to SEQ ID NO:9).

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1, 3, 4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____ , or a portion of any of these nucleotide sequences. A nucleic acid molecule which is complementary to the nucleotide sequence shown in SEQ ID NO:1, 3, 4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, is one which is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO:1, 3, 4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, such that it can hybridize to the nucleotide sequence shown in SEQ ID NO:1, 3, 4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least about 32.2%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more identical to the entire length of the nucleotide sequence shown in SEQ ID NO:1 or 3, or the entire length of the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or a portion of any of these nucleotide sequences.

In still another preferred embodiment, an isolated nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least about 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more identical to the entire length of the nucleotide sequence shown in SEQ ID NO:4, 6, 7, or 9, or the entire length of the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or a portion of any of these nucleotide sequences.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, for example, a fragment which can be used as a probe or primer or a fragment encoding a portion of an FGF-20 protein, *e.g.*, a biologically active portion of an FGF-20 protein. The nucleotide sequence determined from the cloning of the FGF-20 gene allows for the generation of probes and primers designed for use in identifying and/or cloning other FGF-20 family members, as well as FGF-20 homologues from other species. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 consecutive nucleotides of a sense sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, of an anti-sense sequence of SEQ ID NO:1, 3, 4, 6, 7 or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or of a naturally occurring allelic variant or mutant of SEQ ID NO:1, 3, 4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____. In one embodiment, a nucleic acid molecule of the present invention comprises a nucleotide sequence which is greater than 107, 107-200, 200-250, 250-300, 300-350, 350-400, 400-450, 450-500, 500-550, 550-600, 600-650, 650-700, 700-750, 750-800, or more nucleotides in length and hybridizes under stringent hybridization conditions to a nucleic acid molecule of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____. In another embodiment, a nucleic acid molecule of the present invention comprises a nucleotide sequence which is greater than 1156, 1156-1200, 1200-1400, 1400-1600, 1600-1800, 1800-2000, 2000-2200, 2200-2328, 2329, 2329-2350, 2350-2400, 2400-2450, 2450-2500, 2500-2550, 2550-2600, 2600-2650, 2650-2700, or more nucleotides in length and hybridizes under stringent hybridization conditions to a nucleic acid molecule of SEQ ID NO:4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____.

Probes based on the FGF-20 nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress an FGF-20 protein, such as by measuring a level of an FGF-20-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting FGF-20 mRNA levels or determining whether a genomic FGF-20 gene has been mutated or deleted.

A nucleic acid fragment encoding a "biologically active portion of an FGF-20 protein" can be prepared by isolating a portion of the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, which encodes a polypeptide having an FGF-20 biological activity (the biological activities of the FGF-20 proteins are described herein), expressing the encoded portion of the FGF-20 protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the FGF-20 protein.

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO:1, 3, 4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, due to degeneracy of the genetic code and thus encode the same FGF-20 proteins as those encoded by the nucleotide sequence shown in SEQ ID NO:1, 3, 4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NO:2, 5, or 8.

In addition to the FGF-20 nucleotide sequences shown in SEQ ID NO:1, 3, 4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the FGF-20 proteins may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the FGF-20 genes may exist among individuals within a

population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which include an open reading frame encoding an FGF-20 protein, preferably a mammalian FGF-20 protein, and can further include non-coding regulatory sequences, and introns.

Allelic variants of monkey and human FGF-20 include both functional and non-functional FGF-20 proteins. Functional allelic variants are naturally occurring amino acid sequence variants of the monkey or human FGF-20 proteins that maintain the ability to bind an FGF-20 ligand or substrate and/or modulate cell proliferation and/or migration mechanisms. Functional allelic variants will typically contain only conservative substitution of one or more amino acids of SEQ ID NO:2, 5, or 8, or substitution, deletion or insertion of non-critical residues in non-critical regions of the protein.

Non-functional allelic variants are naturally occurring amino acid sequence variants of the monkey or human FGF-20 proteins that do not have the ability to either bind an FGF-20 ligand or substrate and/or modulate cell proliferation and/or migration mechanisms. Non-functional allelic variants will typically contain a non-conservative substitution, a deletion, or insertion or premature truncation of the amino acid sequence of SEQ ID NO:2, 5, or 8, or a substitution, insertion or deletion in critical residues or critical regions.

The present invention further provides non-monkey and non-human orthologues of the monkey and FGF-20 proteins. Orthologues of the monkey and human FGF-20 proteins are proteins that are isolated from non-monkey and non-human organisms and possess the same FGF-20 ligand binding and/or modulation of cell proliferation and/or migration mechanisms of the monkey or human FGF-20 proteins. Orthologues of the monkey or human FGF-20 proteins can readily be identified as comprising an amino acid sequence that is substantially identical to SEQ ID NO:2, 5, or 8.

Moreover, nucleic acid molecules encoding other FGF-20 family members and, thus, which have a nucleotide sequence which differs from the FGF-20 sequences of SEQ ID NO:1, 3, 4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____ are intended to be within the scope of the invention. For example, another FGF-20 cDNA can be identified based on the nucleotide sequence of monkey or human FGF-20. Moreover, nucleic acid

molecules encoding FGF-20 proteins from different species, and which, thus, have a nucleotide sequence which differs from the FGF-20 sequences of SEQ ID NO:1, 3, 4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____ are intended to be within the scope of the invention. For example, a mouse FGF-20 cDNA can be identified based on the nucleotide sequence of a monkey or human FGF-20.

Nucleic acid molecules corresponding to natural allelic variants and homologues of the FGF-20 cDNAs of the invention can be isolated based on their homology to the FGF-20 nucleic acids disclosed herein using the cDNAs disclosed herein, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. Nucleic acid molecules corresponding to natural allelic variants and homologues of the FGF-20 cDNAs of the invention can further be isolated by mapping to the same chromosome or locus as the FGF-20 gene.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15, 20, 25, 30 or more nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____. In other embodiment, the nucleic acid is at least 30, 50, 100, 107, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 1000, 1156, 1200, 1400, 1600, 1800, 2000, 2200, 2329, 2400, or more nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% identical to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85% or 90% identical to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50°C, preferably at 55°C, more preferably at 60°C, and even more preferably at 65°C. Preferably, an isolated nucleic acid molecule of the invention that

hybridizes under stringent conditions to the sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In addition to naturally-occurring allelic variants of the FGF-20 sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NO:1, 3, 4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, thereby leading to changes in the amino acid sequence of the encoded FGF-20 proteins, without altering the functional ability of the FGF-20 proteins. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of FGF-20 (*e.g.*, the sequence of SEQ ID NO:2, 5, or 8) without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the FGF-20 proteins of the present invention, *e.g.*, those present in the fibroblast growth factor domain, are predicted to be particularly unamenable to alteration. Furthermore, additional amino acid residues that are conserved between the FGF-20 proteins of the present invention and other members of the FGF family are not likely to be amenable to alteration.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding FGF-20 proteins that contain changes in amino acid residues that are not essential for activity. Such FGF-20 proteins differ in amino acid sequence from SEQ ID NO:2, 5, or 8, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more identical to SEQ ID NO:2, 5, or 8.

An isolated nucleic acid molecule encoding an FGF-20 protein identical to the protein of SEQ ID NO:2, 5, or 8, can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:1, 3, 4,

6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into SEQ ID NO:1, 3, 4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____ by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in an FGF-20 protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an FGF-20 coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for FGF-20 biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1, 3, 4, 6, 7, or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

In a preferred embodiment, a mutant FGF-20 protein can be assayed for the ability to (1) interact with a non-FGF-20 protein molecule, *e.g.*, an FGF-20 ligand or substrate; (2) activate an FGF-20-dependent signal transduction pathway; or (3) modulate cell proliferation and/or migration mechanisms.

In addition to the nucleic acid molecules encoding FGF-20 proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which

is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire FGF-20 coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding FGF-20. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (*e.g.*, the coding region of monkey FGF-20 corresponds to SEQ ID NO:3 and the coding region of human FGF-20 corresponds to SEQ ID NO:6 or 9). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding FGF-20. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding FGF-20 disclosed herein (*e.g.*, SEQ ID NO:3, 6 or 9), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of FGF-20 mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of FGF-20 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of FGF-20 mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-

carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an FGF-20 protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention include direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett.* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave FGF-20 mRNA transcripts to thereby inhibit translation of FGF-20 mRNA. A ribozyme having specificity for an FGF-20-encoding nucleic acid can be designed based upon the nucleotide sequence of an FGF-20 cDNA disclosed herein (*i.e.*, SEQ ID NO:1, 3, 4, 6, 7 or 9, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an FGF-20-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Patent No. 4,987,071; and Cech *et al.* U.S. Patent No. 5,116,742. Alternatively, FGF-20 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

Alternatively, FGF-20 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the FGF-20 (*e.g.*, the FGF-20 promoter and/or enhancers; *e.g.*, residues 1-325 of SEQ ID NO:4 or residues 1-1070 of SEQ ID NO:7) to form triple helical structures that prevent transcription of the FGF-20 gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15.

In yet another embodiment, the FGF-20 nucleic acid molecules of the present invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup B. *et al.* (1996) *Bioorganic & Medicinal Chemistry* 4 (1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *et al.* Proc. Natl. Acad. Sci. 93: 14670-675.

PNAs of FGF-20 nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of FGF-20 nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (*e.g.*, by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes, (*e.g.*, S1 nucleases (Hyrup B. (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *supra*).

In another embodiment, PNAs of FGF-20 can be modified, (*e.g.*, to enhance their stability or cellular uptake), by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of FGF-20 nucleic acid molecules can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, (*e.g.*, RNAse H and DNA polymerases), to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup B. (1996) *supra*). The synthesis of

PNA-DNA chimeras can be performed as described in Hyrup B. (1996) *supra* and Finn P.J. *et al.* (1996) *Nucleic Acids Res.* 24 (17): 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used as a between the PNA and the 5' end of DNA (Mag, M. *et al.* (1989) *Nucleic Acid Res.* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn P.J. *et al.* (1996) *supra*). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser, K.H. *et al.* (1975) *Bioorganic Med. Chem. Lett.* 5: 1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (See, *e.g.*, Krol *et al.* (1988) *Bio-Techniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, (*e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

II. Isolated FGF-20 Proteins and Anti-FGF-20 Antibodies

One aspect of the invention pertains to isolated FGF-20 proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-FGF-20 antibodies. In one embodiment, native FGF-20 proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, FGF-20 proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, an FGF-20 protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the FGF-20 protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of FGF-20 protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of FGF-20 protein having less than about 30% (by dry weight) of non-FGF-20 protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-FGF-20 protein, still more preferably less than about 10% of non-FGF-20 protein, and most preferably less than about 5% non-FGF-20 protein. When the FGF-20 protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of FGF-20 protein in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of FGF-20 protein having less than about 30% (by dry weight) of chemical precursors or non-FGF-20 chemicals, more preferably less than about 20% chemical precursors or non-FGF-20 chemicals, still more preferably less than about 10% chemical precursors or non-FGF-20 chemicals, and most preferably less than about 5% chemical precursors or non-FGF-20 chemicals.

As used herein, a "biologically active portion" of an FGF-20 protein includes a fragment of an FGF-20 protein which participates in an interaction between an FGF-20 molecule and a non-FGF-20 molecule. Biologically active portions of an FGF-20 protein include peptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the FGF-20 protein, *e.g.*, the amino acid sequence shown in SEQ ID NO:2, 5, or 8, which include less amino acids than the full length FGF-20 proteins, and exhibit at least one activity of an FGF-20 protein.

Typically, biologically active portions comprise a domain or motif with at least one activity of the FGF-20 protein, *e.g.*, modulating cell proliferation mechanisms. A biologically active portion of an FGF-20 protein can be a polypeptide which is, for example, 10, 25, 50, 100, 177, or more amino acids in length. Biologically active portions of an FGF-20 protein can be used as targets for developing agents which modulate an FGF-20 mediated activity, *e.g.*, a cell proliferation mechanism.

In one embodiment, a biologically active portion of an FGF-20 protein comprises at least one fibroblast growth factor domain, and/or at least one beta trefoil structure. It is to be understood that a preferred biologically active portion of an FGF-20 protein of the present invention may contain at least one fibroblast growth factor domain. Another preferred biologically active portion of an FGF-20 protein may contain a beta trefoil structure. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native FGF-20 protein.

In a preferred embodiment, the FGF-20 protein has an amino acid sequence shown in SEQ ID NO:2, 5, or 8. In other embodiments, the FGF-20 protein is substantially identical to SEQ ID NO:2, 5, or 8, and retains the functional activity of the protein of SEQ ID NO:2, 5, or 8, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail in subsection I above. Accordingly, in another embodiment, the FGF-20 protein is a protein which comprises an amino acid sequence at least about 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more identical to SEQ ID NO:2, 5, or 8.

To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-identical sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, or 90% of the length of the reference sequence (*e.g.*, when aligning a second sequence to the FGF-20 amino acid sequence of SEQ ID NO:2 having 177 amino acid residues, at least 53, preferably at least 71, more preferably at least 89, even

more preferably at least 106, and even more preferably at least 124, 142 or 159 amino acid residues are aligned). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Meyers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to FGF-20 nucleic acid molecules of the invention. BLAST protein

searches can be performed with the XBLAST program, score = 100, wordlength = 3 to obtain amino acid sequences homologous to FGF-20 protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov.

The invention also provides FGF-20 chimeric or fusion proteins. As used herein, an FGF-20 "chimeric protein" or "fusion protein" comprises an FGF-20 polypeptide operatively linked to a non-FGF-20 polypeptide. An "FGF-20 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to FGF-20, whereas a "non-FGF-20 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the FGF-20 protein, *e.g.*, a protein which is different from the FGF-20 protein and which is derived from the same or a different organism. Within an FGF-20 fusion protein the FGF-20 polypeptide can correspond to all or a portion of an FGF-20 protein. In a preferred embodiment, an FGF-20 fusion protein comprises at least one biologically active portion of an FGF-20 protein. In another preferred embodiment, an FGF-20 fusion protein comprises at least two biologically active portions of an FGF-20 protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the FGF-20 polypeptide and the non-FGF-20 polypeptide are fused in-frame to each other. The non-FGF-20 polypeptide can be fused to the N-terminus or C-terminus of the FGF-20 polypeptide.

For example, in one embodiment, the fusion protein is a GST-FGF-20 fusion protein in which the FGF-20 sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant FGF-20.

In another embodiment, the fusion protein is an FGF-20 protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of FGF-20 can be increased through use of a heterologous signal sequence.

The FGF-20 fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. The FGF-20 fusion proteins can be used to affect the bioavailability of an FGF-20 substrate. Use of FGF-20 fusion proteins may be useful therapeutically for the treatment of disorders caused by,

for example, (i) aberrant modification or mutation of a gene encoding an FGF-20 protein; (ii) mis-regulation of the FGF-20 gene; and (iii) aberrant post-translational modification of an FGF-20 protein.

Moreover, the FGF-20-fusion proteins of the invention can be used as immunogens to produce anti-FGF-20 antibodies in a subject, to purify FGF-20 ligands and in screening assays to identify molecules which inhibit the interaction of FGF-20 with an FGF-20 substrate.

Preferably, an FGF-20 chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel *et al.* John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). An FGF-20-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the FGF-20 protein.

The present invention also pertains to variants of the FGF-20 proteins which function as either FGF-20 agonists (mimetics) or as FGF-20 antagonists. Variants of the FGF-20 proteins can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of an FGF-20 protein. An agonist of the FGF-20 proteins can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of an FGF-20 protein. An antagonist of an FGF-20 protein can inhibit one or more of the activities of the naturally occurring form of the FGF-20 protein by, for example, competitively modulating an FGF-20-mediated activity of an FGF-20 protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function.

In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the FGF-20 protein.

In one embodiment, variants of an FGF-20 protein which function as either FGF-20 agonists (mimetics) or as FGF-20 antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of an FGF-20 protein for FGF-20 protein agonist or antagonist activity. In one embodiment, a variegated library of FGF-20 variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of FGF-20 variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential FGF-20 sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of FGF-20 sequences therein. There are a variety of methods which can be used to produce libraries of potential FGF-20 variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential FGF-20 sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu. Rev. Biochem.* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucleic Acid Res.* 11:477.

In addition, libraries of fragments of an FGF-20 protein coding sequence can be used to generate a variegated population of FGF-20 fragments for screening and subsequent selection of variants of an FGF-20 protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an FGF-20 coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into

an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the FGF-20 protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of FGF-20 proteins. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recrusive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify FGF-20 variants (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave *et al.* (1993) *Protein Engineering* 6(3):327-331).

In one embodiment, cell based assays can be exploited to analyze a variegated FGF-20 library. For example, a library of expression vectors can be transfected into a cell line, *e.g.*, an endothelial cell line, which ordinarily responds to FGF-20 in a particular FGF-20 substrate-dependent manner. The transfected cells are then contacted with FGF-20 and the effect of expression of the mutant on signaling by the FGF-20 substrate can be detected, *e.g.*, by monitoring intracellular calcium, IP3, or diacylglycerol concentration, phosphorylation profile of intracellular proteins, cell proliferation and/or migration, or the activity of an FGF-20-regulated transcription factor. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of signaling by the FGF-20 substrate, and the individual clones further characterized.

An isolated FGF-20 protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind FGF-20 using standard techniques for polyclonal and monoclonal antibody preparation. A full-length FGF-20 protein can be used or, alternatively, the invention provides antigenic peptide fragments of FGF-20 for

use as immunogens. The antigenic peptide of FGF-20 comprises at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:2, 5, or 8 and encompasses an epitope of FGF-20 such that an antibody raised against the peptide forms a specific immune complex with FGF-20. Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

Preferred epitopes encompassed by the antigenic peptide are regions of FGF-20 that are located on the surface of the protein, *e.g.*, hydrophilic regions, as well as regions with high antigenicity (see, for example, Figures 2 and 13).

An FGF-20 immunogen typically is used to prepare antibodies by immunizing a suitable subject, (*e.g.*, rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed FGF-20 protein or a chemically synthesized FGF-20 polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic FGF-20 preparation induces a polyclonal anti-FGF-20 antibody response.

Accordingly, another aspect of the invention pertains to anti-FGF-20 antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site which specifically binds (immunoreacts with) an antigen, such as FGF-20. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind FGF-20. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of FGF-20. A monoclonal antibody composition thus typically displays a single binding affinity for a particular FGF-20 protein with which it immunoreacts.

Polyclonal anti-FGF-20 antibodies can be prepared as described above by immunizing a suitable subject with an FGF-20 immunogen. The anti-FGF-20 antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized FGF-20. If desired, the antibody molecules directed against FGF-20 can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, *e.g.*, when the anti-FGF-20 antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497) (see also, Brown *et al.* (1981) *J. Immunol.* 127:539-46; Brown *et al.* (1980) *J. Biol. Chem* .255:4980-83; Yeh *et al.* (1976) *Proc. Natl. Acad. Sci. USA* 76:2927-31; and Yeh *et al.* (1982) *Int. J. Cancer* 29:269-75), the more recent human B cell hybridoma technique (Kozbor *et al.* (1983) *Immunol Today* 4:72), the EBV-hybridoma technique (Cole *et al.* (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing monoclonal antibody hybridomas is well known (see generally R. H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); E. A. Lerner (1981) *Yale J. Biol. Med.,* 54:387-402; M. L. Gefter *et al.* (1977) *Somatic Cell Genet.* 3:231-36). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with an FGF-20 immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds FGF-20.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-FGF-20 monoclonal antibody (see, *e.g.*, G. Galfre *et al.* (1977) *Nature* 266:55052; Gefter *et al.* *Somatic Cell Genet.*, cited *supra*; Lerner, *Yale J. Biol. Med.*, cited *supra*; Kenneth, *Monoclonal Antibodies*, cited *supra*). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (*e.g.*, a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made

by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line. Preferred immortal cell lines are mouse myeloma cell lines that are sensitive to culture medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques, *e.g.*, the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused myeloma cells (unfused splenocytes die after several days because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind FGF-20, *e.g.*, using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-FGF-20 antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display library) with FGF-20 to thereby isolate immunoglobulin library members that bind FGF-20. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP™ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, Ladner *et al.* U.S. Patent No. 5,223,409; Kang *et al.* PCT International Publication No. WO 92/18619; Dower *et al.* PCT International Publication No. WO 91/17271; Winter *et al.* PCT International Publication WO 92/20791; Markland *et al.* PCT International Publication No. WO 92/15679; Breitling *et al.* PCT International Publication WO 93/01288; McCafferty *et al.* PCT International Publication No. WO 92/01047; Garrard *et al.* PCT International Publication No. WO 92/09690; Ladner *et al.* PCT International Publication No. WO 90/02809; Fuchs *et al.* (1991) *Bio/Technology* 9:1370-1372; Hay *et al.* (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse *et al.* (1989) *Science* 246:1275-1281; Griffiths *et al.* (1993) *EMBO J* 12:725-734; Hawkins *et al.* (1992) *J. Mol. Biol.* 226:889-896; Clarkson *et al.* (1991) *Nature* 352:624-628; Gram *et al.* (1992) *Proc. Natl. Acad.*

Sci. USA 89:3576-3580; Garrad *et al.* (1991) *Bio/Technology* 9:1373-1377; Hoogenboom *et al.* (1991) *Nuc. Acid Res.* 19:4133-4137; Barbas *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:7978-7982; and McCafferty *et al. Nature* (1990) 348:552-554.

Additionally, recombinant anti-FGF-20 antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson *et al.* International Application No. PCT/US86/02269; Akira, *et al.* European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison *et al.* European Patent Application 173,494; Neuberger *et al.* PCT International Publication No. WO 86/01533; Cabilly *et al.* U.S. Patent No. 4,816,567; Cabilly *et al.* European Patent Application 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu *et al.* (1987) *J. Immunol.* 139:3521-3526; Sun *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura *et al.* (1987) *Canc. Res.* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; and Shaw *et al.* (1988) *J. Natl. Cancer Inst.* 80:1553-1559); Morrison, S. L. (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *BioTechniques* 4:214; Winter U.S. Patent 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525; Verhoeyan *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J. Immunol.* 141:4053-4060.

An anti-FGF-20 antibody (*e.g.*, monoclonal antibody) can be used to isolate FGF-20 by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-FGF-20 antibody can facilitate the purification of natural FGF-20 from cells and of recombinantly produced FGF-20 expressed in host cells. Moreover, an anti-FGF-20 antibody can be used to detect FGF-20 protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the FGF-20 protein. Anti-FGF-20 antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive

materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

III. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an FGF-20 protein (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory

sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cells and those which direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, FGF-20 proteins, mutant forms of FGF-20 proteins, fusion proteins, and the like).

The recombinant expression vectors of the invention can be designed for expression of FGF-20 proteins in prokaryotic or eukaryotic cells. For example, FGF-20 proteins can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein;

2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Purified fusion proteins can be utilized in FGF-20 activity assays, (*e.g.*, direct assays or competitive assays described in detail below), or to generate antibodies specific for FGF-20 proteins, for example. In a preferred embodiment, an FGF-20 fusion protein expressed in a retroviral expression vector of the present invention can be utilized to infect bone marrow cells which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (*e.g.*, six (6) weeks).

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an

expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the FGF-20 expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerivisae* include pYepSec1 (Baldari, *et al.*, (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and picZ (InVitrogen Corp, San Diego, CA).

Alternatively, FGF-20 proteins can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, Sf 9 cells) include the pAc series (Smith *et al.* (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell*

33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The expression characteristics of an endogenous FGF-20 gene within a cell line or microorganism may be modified by inserting a heterologous DNA regulatory element into the genome of a stable cell line or cloned microorganism such that the inserted regulatory element is operatively linked with the endogenous FGF-20 gene. For example, an endogenous FGF-20 gene which is normally "transcriptionally silent", *i.e.*, a FGF-20 gene which is normally not expressed, or is expressed only at very low levels in a cell line or microorganism, may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell line or microorganism. Alternatively, a transcriptionally silent, endogenous FGF-20 gene may be activated by insertion of a promiscuous regulatory element that works across cell types.

A heterologous regulatory element may be inserted into a stable cell line or cloned microorganism, such that it is operatively linked with an endogenous FGF-20 gene, using techniques, such as targeted homologous recombination, which are well known to those of skill in the art, and described, *e.g.*, in Chappel, U.S. Patent No. 5,272,071; PCT publication No. WO 91/06667, published May 16, 1991.

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to FGF-20 mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which

direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. *et al.*, Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which an FGF-20 nucleic acid molecule of the invention is introduced, *e.g.*, an FGF-20 nucleic acid molecule within a recombinant expression vector or an FGF-20 nucleic acid molecule containing sequences which allow it to homologously recombine into a specific site of the host cell's genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, an FGF-20 protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding an FGF-20 protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) an FGF-20 protein. Accordingly, the invention further provides methods for producing an FGF-20 protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of the invention (into which a recombinant expression vector encoding an FGF-20 protein has been introduced) in a suitable medium such that an FGF-20 protein is produced. In another embodiment, the method further comprises isolating an FGF-20 protein from the medium or the host cell.

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which FGF-20-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous FGF-20 sequences have been introduced into their genome or homologous recombinant animals in which endogenous FGF-20 sequences have been altered. Such animals are useful for studying the function and/or activity of an FGF-20 and for identifying and/or evaluating modulators of FGF-20 activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops

and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous FGF-20 gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing an FGF-20-encoding nucleic acid into the male pronuclei of a fertilized oocyte, *e.g.*, by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The FGF-20 cDNA sequence of SEQ ID NO:1, 4, or 7 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of a human FGF-20 gene, such as a mouse or rat FGF-20 gene, can be used as a transgene. Alternatively, an FGF-20 gene homologue, such as another FGF-20 family member, can be isolated based on hybridization to the FGF-20 cDNA sequences of SEQ ID NO:1, 3, 4, 6, 7, or 9, or the DNA insert of the plasmid deposited with ATCC as Accession Number _____ (described further in subsection I above) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to an FGF-20 transgene to direct expression of an FGF-20 protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of an FGF-20 transgene in its genome and/or expression of FGF-20 mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding an

FGF-20 protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of an FGF-20 gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the FGF-20 gene. The FGF-20 gene can be a monkey gene (*e.g.*, the cDNA of SEQ ID NO:3) or a human gene (*e.g.*, the cDNA of SEQ ID NO:6 or 9), but more preferably, is a non-human homologue of a monkey or human FGF-20 gene (*e.g.*, a cDNA isolated by stringent hybridization with the nucleotide sequence of SEQ ID NO:1, 4, or 7). For example, a mouse FGF-20 gene can be used to construct a homologous recombination nucleic acid molecule, *e.g.*, a vector, suitable for altering an endogenous FGF-20 gene in the mouse genome. In a preferred embodiment, the homologous recombination nucleic acid molecule is designed such that, upon homologous recombination, the endogenous FGF-20 gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the homologous recombination nucleic acid molecule can be designed such that, upon homologous recombination, the endogenous FGF-20 gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous FGF-20 protein). In the homologous recombination nucleic acid molecule, the altered portion of the FGF-20 gene is flanked at its 5' and 3' ends by additional nucleic acid sequence of the FGF-20 gene to allow for homologous recombination to occur between the exogenous FGF-20 gene carried by the homologous recombination nucleic acid molecule and an endogenous FGF-20 gene in a cell, *e.g.*, an embryonic stem cell. The additional flanking FGF-20 nucleic acid sequence is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the homologous recombination nucleic acid molecule (see, *e.g.*, Thomas, K.R. and Capecchi, M. R. (1987) *Cell* 51:503 for a description of homologous recombination vectors). The homologous recombination nucleic acid molecule is introduced into a cell, *e.g.*, an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced FGF-20 gene has homologously recombined with the endogenous FGF-20 gene are selected (see *e.g.*, Li, E. *et al.* (1992) *Cell* 69:915). The selected cells can then injected

into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras (see *e.g.*, Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination nucleic acid molecules, *e.g.*, vectors, or homologous recombinant animals are described further in Bradley, A. (1991) *Current Opinion in Biotechnology* 2:823-829 and in PCT International Publication Nos.: WO 90/11354 by Le Mouellec *et al.*; WO 91/01140 by Smithies *et al.*; WO 92/0968 by Zijlstra *et al.*; and WO 93/04169 by Berns *et al.*

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, *e.g.*, Lakso *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman *et al.* (1991) *Science* 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al.* (1997) *Nature* 385:810-813 and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, *e.g.*, a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_O phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to

pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, *e.g.*, the somatic cell, is isolated.

IV. Pharmaceutical Compositions

The FGF-20 nucleic acid molecules, fragments of FGF-20 proteins, and anti-FGF-20 antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a fragment of an FGF-20 protein or an anti-FGF-20 antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate,

polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Compounds which exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a

circulating plasma concentration range that includes the IC50 (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

As defined herein, a therapeutically effective amount of protein or polypeptide (*i.e.*, an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The skilled artisan will appreciate that certain factors may influence the dosage required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments.

In a preferred example, a subject is treated with antibody, protein, or polypeptide in the range of between about 0.1 to 20 mg/kg body weight, one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. It will also be appreciated that the effective dosage of antibody, protein, or polypeptide used for treatment may increase or decrease over the course of a particular treatment. Changes in dosage may result and become apparent from the results of diagnostic assays as described herein.

The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics, amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (*i.e*,. including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than

about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds. It is understood that appropriate doses of small molecule agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of the small molecule will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the small molecule to have upon the nucleic acid or polypeptide of the invention.

Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (*e.g.*, about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. Such appropriate doses may be determined using the assays described herein. When one or more of these small molecules is to be administered to an animal (*e.g.*, a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

Further, an antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs

or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (*e.g.*, mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclothosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (*e.g.*, daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (*e.g.*, dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (*e.g.*, vincristine and vinblastine).

The conjugates of the invention can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, .alpha.-interferon, .beta.-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophase colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

Techniques for conjugating such therapeutic moiety to antibodies are well known, see, *e.g.*, Arnon et al., "Monoclonal Antibodies For Immunotargeting Of Drugs In Cancer Therapy", in Monoclonal Antibodies And Cancer Therapy, Reisfeld et al. (eds.), pp. 243-56 (Alan R. Liss, Inc. 1985); Hellstrom et al., "Antibodies For Drug Delivery", in Controlled Drug Delivery (2nd Ed.), Robinson et al. (eds.), pp. 623-53 (Marcel Dekker, Inc. 1987); Thorpe, "Antibody Carriers Of Cytotoxic Agents In Cancer Therapy: A Review", in Monoclonal Antibodies '84: Biological And Clinical Applications, Pinchera et al. (eds.), pp. 475-506 (1985); "Analysis, Results, And Future Prospective Of The Therapeutic Use Of Radiolabeled Antibody In Cancer Therapy", in Monoclonal Antibodies For Cancer Detection And Therapy, Baldwin et al. (eds.), pp. 303-16 (Academic Press 1985), and Thorpe et al., "The Preparation And Cytotoxic Properties Of Antibody-Toxin Conjugates", Immunol. Rev., 62:119-58 (1982).

Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent 5,328,470) or by stereotactic injection (see *e.g.*, Chen *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

V. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (*e.g.*, therapeutic and prophylactic). As described herein, an FGF-20 protein of the invention has one or more of the following activities: (1) it interacts with a non-FGF-20 protein molecule, *e.g.*, a FGF-20 substrate, such as a FGF receptor or heparan sulfate proteoglycan; (2) it activates an FGF-20-dependent signal transduction pathway; and (3) it modulates cell proliferation and/or migration mechanisms, and, thus, can be used to, for example, (1) modulate the interaction with a non-FGF-20 protein molecule; (2) to activate an FGF-20-dependent signal transduction pathway; and (3) to modulate cell proliferation and/or migration mechanisms.

The isolated nucleic acid molecules of the invention can be used, for example, to express FGF-20 protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect FGF-20 mRNA (*e.g.*, in a biological sample) or a genetic alteration in an FGF-20 gene, and to modulate FGF-20 activity, as described further

below. The FGF-20 proteins can be used to treat disorders characterized by insufficient or excessive production of an FGF-20 substrate or production of FGF-20 inhibitors. In addition, the FGF-20 proteins can be used to screen for naturally occurring FGF-20 substrates, to screen for drugs or compounds which modulate FGF-20 activity, as well as to treat disorders characterized by insufficient or excessive production of FGF-20 protein or production of FGF-20 protein forms which have decreased, aberrant or unwanted activity compared to FGF-20 wild type protein (*e.g.*, proliferative disorders, neurodegenerative disorders, cardiovascular disorders, or pain disorders). Moreover, the anti-FGF-20 antibodies of the invention can be used to detect and isolate FGF-20 proteins, regulate the bioavailability of FGF-20 proteins, and modulate FGF-20 activity.

A. Screening Assays:

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) which bind to FGF-20 proteins, have a stimulatory or inhibitory effect on, for example, FGF-20 expression or FGF-20 activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of FGF-20 substrate.

In one embodiment, the invention provides assays for screening candidate or test compounds which are substrates of an FGF-20 protein or polypeptide or biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of an FGF-20 protein or polypeptide or biologically active portion thereof. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994). *J. Med. Chem.* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP '409), plasmids (Cull *et al.* (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390); (Devlin (1990) *Science* 249:404-406); (Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382); (Felici (1991) *J. Mol. Biol.* 222:301-310); (Ladner *supra.*).

In one embodiment, an assay is a cell-based assay in which a cell which expresses an FGF-20 protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate FGF-20 activity is determined. Determining the ability of the test compound to modulate FGF-20 activity can be accomplished by monitoring, for example, intracellular calcium, IP3, or diacylglycerol concentration, phosphorylation profile of intracellular proteins, cell proliferation and/or migration, or the activity of an FGF-20-regulated transcription factor. The cell, for example, can be of mammalian origin, *e.g.*, an endothelial cell.

The ability of the test compound to modulate FGF-20 binding to a substrate or to bind to FGF-20 can also be determined. Determining the ability of the test compound to modulate FGF-20 binding to a substrate can be accomplished, for example, by coupling the FGF-20 substrate with a radioisotope or enzymatic label such that binding of the FGF-20 substrate to FGF-20 can be determined by detecting the labeled FGF-20 substrate in a complex. Alternatively, FGF-20 could be coupled with a radioisotope or enzymatic label to monitor the ability of a test compound to modulate FGF-20 binding to a FGF-20 substrate in a complex. Determining the ability of the test compound to bind FGF-20 can be accomplished, for example, by coupling the compound with a radioisotope or enzymatic label such that binding of the compound to FGF-20 can be determined by detecting the labeled FGF-20 compound in a complex. For example,

compounds (*e.g.*, FGF-20 substrates) can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

It is also within the scope of this invention to determine the ability of a compound (*e.g.*, an FGF-20 substrate) to interact with FGF-20 without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of a compound with FGF-20 without the labeling of either the compound or the FGF-20. McConnell, H. M. *et al.* (1992) *Science* 257:1906-1912. As used herein, a "microphysiometer" (*e.g.*, Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a compound and FGF-20.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing an FGF-20 target molecule (*e.g.*, an FGF-20 substrate) with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the FGF-20 target molecule. Determining the ability of the test compound to modulate the activity of an FGF-20 target molecule can be accomplished, for example, by determining the ability of the FGF-20 protein to bind to or interact with the FGF-20 target molecule.

Determining the ability of the FGF-20 protein or a biologically active fragment thereof, to bind to or interact with an FGF-20 target molecule can be accomplished by one of the methods described above for determining direct binding. In a preferred embodiment, determining the ability of the FGF-20 protein to bind to or interact with an FGF-20 target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.*, intracellular Ca^{2+}, diacylglycerol, IP_3, and the like), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising a target-

responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a target-regulated cellular response.

In yet another embodiment, an assay of the present invention is a cell-free assay in which an FGF-20 protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the FGF-20 protein or biologically active portion thereof is determined. Preferred biologically active portions of the FGF-20 proteins to be used in assays of the present invention include fragments which participate in interactions with non-FGF-20 molecules, *e.g.*, fragments with high surface probability scores (see, for example, Figures 2 and 13). Binding of the test compound to the FGF-20 protein can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the FGF-20 protein or biologically active portion thereof with a known compound which binds FGF-20 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an FGF-20 protein, wherein determining the ability of the test compound to interact with an FGF-20 protein comprises determining the ability of the test compound to preferentially bind to FGF-20 or biologically active portion thereof as compared to the known compound.

In another embodiment, the assay is a cell-free assay in which an FGF-20 protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the FGF-20 protein or biologically active portion thereof is determined. Determining the ability of the test compound to modulate the activity of an FGF-20 protein can be accomplished, for example, by determining the ability of the FGF-20 protein to bind to an FGF-20 target molecule by one of the methods described above for determining direct binding. Determining the ability of the FGF-20 protein to bind to an FGF-20 target molecule can also be accomplished using a technology such as real-time Biomolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (*e.g.*, BIAcore). Changes in the optical phenomenon of surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

In an alternative embodiment, determining the ability of the test compound to modulate the activity of an FGF-20 protein can be accomplished by determining the ability of the FGF-20 protein to further modulate the activity of a downstream effector of an FGF-20 target molecule. For example, the activity of the effector molecule on an appropriate target can be determined or the binding of the effector to an appropriate target can be determined as previously described.

In yet another embodiment, the cell-free assay involves contacting an FGF-20 protein or biologically active portion thereof with a known compound which binds the FGF-20 protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the FGF-20 protein, wherein determining the ability of the test compound to interact with the FGF-20 protein comprises determining the ability of the FGF-20 protein to preferentially bind to or modulate the activity of an FGF-20 target molecule.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either FGF-20 or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to an FGF-20 protein, or interaction of an FGF-20 protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/ FGF-20 fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or FGF-20 protein, and the mixture incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively,

the complexes can be dissociated from the matrix, and the level of FGF-20 binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either an FGF-20 protein or an FGF-20 target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated FGF-20 protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with FGF-20 protein or target molecules but which do not interfere with binding of the FGF-20 protein to its target molecule can be derivatized to the wells of the plate, and unbound target or FGF-20 protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the FGF-20 protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the FGF-20 protein or target molecule.

In another embodiment, modulators of FGF-20 expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of FGF-20 mRNA or protein in the cell is determined. The level of expression of FGF-20 mRNA or protein in the presence of the candidate compound is compared to the level of expression of FGF-20 mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of FGF-20 expression based on this comparison. For example, when expression of FGF-20 mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of FGF-20 mRNA or protein expression. Alternatively, when expression of FGF-20 mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of FGF-20 mRNA or protein expression. The level of FGF-20 mRNA or protein expression in the cells can be determined by methods described herein for detecting FGF-20 mRNA or protein.

In yet another aspect of the invention, the FGF-20 proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, *e.g.*, U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with FGF-20 ("FGF-20-binding proteins" or "FGF-20-bp") and are involved in FGF-20 activity. Such FGF-20-binding proteins are also likely to be involved in the propagation of signals by the FGF-20 proteins or FGF-20 targets as, for example, downstream elements of an FGF-20-mediated signaling pathway. Alternatively, such FGF-20-binding proteins are likely to be FGF-20 inhibitors.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for an FGF-20 protein is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo,* forming an FGF-20-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the FGF-20 protein.

In another aspect, the invention pertains to a combination of two or more of the assays described herein. For example, a modulating agent can be identified using a cell-based or a cell free assay, and the ability of the agent to modulate the activity of an FGF-20 protein can be confirmed *in vivo*, *e.g.*, in an animal such as an animal model for cellular transformation and/or tumorigenesis.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an

agent identified as described herein (*e.g.*, an FGF-20 modulating agent, an antisense FGF-20 nucleic acid molecule, an FGF-20-specific antibody, or an FGF-20-binding partner) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

B. Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

1. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the FGF-20 nucleotide sequences, described herein, can be used to map the location of the FGF-20 genes on a chromosome. The mapping of the FGF-20 sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, FGF-20 genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the FGF-20 nucleotide sequences. Computer analysis of the FGF-20 sequences can be used to predict primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the FGF-20 sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (*e.g.*, human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but human cells can, the one human chromosome that contains the gene encoding the needed enzyme, will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. (D'Eustachio P. *et al.* (1983) *Science* 220:919-924). Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the FGF-20 nucleotide sequences to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map an FGF-20 sequence to its chromosome include *in situ* hybridization (described in Fan, Y. *et al.* (1990) *Proc. Natl. Acad. Sci. USA*, 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical such as colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this

technique, see Verma *et al.*, Human Chromosomes: A Manual of Basic Techniques (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland, J. *et al.* (1987) *Nature,* 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the FGF-20 gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

2. Tissue Typing

The FGF-20 sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations

of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the FGF-20 nucleotide sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The FGF-20 nucleotide sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1, 4 or 7 can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:3, 6 or 9 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from FGF-20 nucleotide sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

3. Use of Partial FGF-20 Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, *e.g.*, hair or skin, or body fluids, *e.g.*, blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (*i.e.* another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:1, 4, or 7 are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the FGF-20 nucleotide sequences or portions thereof, *e.g.*, fragments derived from the noncoding regions of SEQ ID NO:1, 4, or 7, having a length of at least 20 bases, preferably at least 30 bases.

The FGF-20 nucleotide sequences described herein can further be used to provide polynucleotide reagents, *e.g.*, labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, *e.g.*, brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such FGF-20 probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, *e.g.*, FGF-20 primers or probes can be used to screen tissue culture for contamination (*i.e.* screen for the presence of a mixture of different types of cells in a culture).

C. Predictive Medicine:

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining FGF-20 protein and/or nucleic acid expression as well as FGF-20 activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant or unwanted FGF-20 expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with FGF-20 protein, nucleic acid expression or activity. For example, mutations in an FGF-20 gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with FGF-20 protein, nucleic acid expression or activity.

Another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of FGF-20 in clinical trials.

These and other agents are described in further detail in the following sections.

1. Diagnostic Assays

An exemplary method for detecting the presence or absence of FGF-20 protein or nucleic acid in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting FGF-20 protein or nucleic acid (*e.g.*, mRNA, or genomic DNA) that encodes FGF-20 protein such that the presence of FGF-20 protein or nucleic acid is detected in the biological sample. A preferred agent for detecting FGF-20 mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to FGF-20 mRNA or genomic DNA. The nucleic acid probe can be, for example, the FGF-20 nucleic acid set forth in SEQ ID NO:1, 3, 4, 6, 7, or 9, or the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize

under stringent conditions to FGF-20 mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

A preferred agent for detecting FGF-20 protein is an antibody capable of binding to FGF-20 protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')2) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect FGF-20 mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of FGF-20 mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of FGF-20 protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of FGF-20 genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of FGF-20 protein include introducing into a subject a labeled anti-FGF-20 antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a serum sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting FGF-20 protein, mRNA, or genomic DNA, such that the presence of FGF-20 protein, mRNA or genomic DNA is detected in the

biological sample, and comparing the presence of FGF-20 protein, mRNA or genomic DNA in the control sample with the presence of FGF-20 protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of FGF-20 in a biological sample. For example, the kit can comprise a labeled compound or agent capable of detecting FGF-20 protein or mRNA in a biological sample; means for determining the amount of FGF-20 in the sample; and means for comparing the amount of FGF-20 in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect FGF-20 protein or nucleic acid.

2. Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant or unwanted FGF-20 expression or activity. As used herein, the term "aberrant" includes an FGF-20 expression or activity which deviates from the wild type FGF-20 expression or activity. Aberrant expression or activity includes increased or decreased expression or activity, as well as expression or activity which does not follow the wild type developmental pattern of expression or the subcellular pattern of expression. For example, aberrant FGF-20 expression or activity is intended to include the cases in which a mutation in the FGF-20 gene causes the FGF-20 gene to be under-expressed or over-expressed and situations in which such mutations result in a non-functional FGF-20 protein or a protein which does not function in a wild-type fashion, *e.g.*, a protein which does not interact with an FGF-20 substrate, *e.g.*, a FGF receptor or heparan sulfate proteoglycan, or one which interacts with a non-FGF-20 substrate, *e.g.* a non-FGF receptor or heparan sulfate proteoglycan. As used herein, the term "unwanted" includes an unwanted phenomenon involved in a biological response such as cellular proliferation. For example, the term unwanted includes an FGF-20 expression or activity which is undesirable in a subject.

The assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with a misregulation in FGF-20 protein activity or nucleic acid

expression, such as a proliferative disorder, a differentiative disorder, a pain disorder, or a cardiovascular disorder. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disorder associated with a misregulation in FGF-20 protein activity or nucleic acid expression, such as a proliferative disorder. Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant or unwanted FGF-20 expression or activity in which a test sample is obtained from a subject and FGF-20 protein or nucleic acid (*e.g.*, mRNA or genomic DNA) is detected, wherein the presence of FGF-20 protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant or unwanted FGF-20 expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant or unwanted FGF-20 expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a proliferative disorder, a differentiative disorder, or a pain disorder. Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant or unwanted FGF-20 expression or activity in which a test sample is obtained and FGF-20 protein or nucleic acid expression or activity is detected (*e.g.*, wherein the abundance of FGF-20 protein or nucleic acid expression or activity is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant or unwanted FGF-20 expression or activity).

The methods of the invention can also be used to detect genetic alterations in an FGF-20 gene, thereby determining if a subject with the altered gene is at risk for a disorder characterized by misregulation in FGF-20 protein activity or nucleic acid expression, such as a proliferative disorder. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding an FGF-20-protein, or the mis-expression of the FGF-20 gene. For

example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from an FGF-20 gene; 2) an addition of one or more nucleotides to an FGF-20 gene; 3) a substitution of one or more nucleotides of an FGF-20 gene, 4) a chromosomal rearrangement of an FGF-20 gene; 5) an alteration in the level of a messenger RNA transcript of an FGF-20 gene, 6) aberrant modification of an FGF-20 gene, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of an FGF-20 gene, 8) a non-wild type level of an FGF-20-protein, 9) allelic loss of an FGF-20 gene, and 10) inappropriate post-translational modification of an FGF-20-protein. As described herein, there are a large number of assays known in the art which can be used for detecting alterations in an FGF-20 gene. A preferred biological sample is a tissue or serum sample isolated by conventional means from a subject.

In certain embodiments, detection of the alteration involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, *e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, *e.g.*, Landegran *et al.* (1988) *Science* 241:1077-1080; and Nakazawa *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:360-364), the latter of which can be particularly useful for detecting point mutations in the FGF-20-gene (see Abravaya *et al.* (1995) *Nucleic Acids Res* .23:675-682). This method can include the steps of collecting a sample of cells from a subject, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to an FGF-20 gene under conditions such that hybridization and amplification of the FGF-20-gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli, J.C. *et al.*, (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, D.Y. *et al.*, (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi, P.M. *et al.* (1988) *Bio-Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified

molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in an FGF-20 gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in FGF-20 can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin, M.T. *et al.* (1996) *Human Mutation* 7: 244-255; Kozal, M.J. *et al.* (1996) *Nature Medicine* 2: 753-759). For example, genetic mutations in FGF-20 can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, M.T. *et al. supra.* Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the FGF-20 gene and detect mutations by comparing the sequence of the sample FGF-20 with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxam and Gilbert ((1977) *Proc. Natl. Acad. Sci. USA* 74:560) or Sanger ((1977) *Proc. Natl. Acad. Sci. USA* 74:5463). It is also contemplated

that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publication No. WO 94/16101; Cohen *et al.* (1996) *Adv. Chromatogr.* 36:127-162; and Griffin *et al.* (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

Other methods for detecting mutations in the FGF-20 gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers *et al.* (1985) *Science* 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type FGF-20 sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton *et al.* (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba *et al.* (1992) *Methods Enzymol.* 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in FGF-20 cDNAs obtained from samples of cells. For example, the mutY enzyme of E. coli cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on an FGF-20 sequence, *e.g.*, a wild-type FGF-20 sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme,

and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in FGF-20 genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (orita *et al.* (1989) *Proc Natl. Acad. Sci USA*: 86:2766, see also Cotton (1993) *Mutat. Res.* 285:125-144; and Hayashi (1992) *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control FGF-20 nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet* 7:5).

In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al.* (1989) *Proc. Natl Acad. Sci USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different

mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini *et al.* (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.,* in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving an FGF-20 gene.

Furthermore, any cell type or tissue in which FGF-20 is expressed may be utilized in the prognostic assays described herein.

3. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs) on the expression or activity of an FGF-20 protein (*e.g.*, the modulation of cell proliferation and/or migration) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase FGF-20 gene expression, protein levels, or upregulate FGF-20 activity, can be monitored in clinical trials of subjects exhibiting decreased FGF-20 gene expression, protein levels, or downregulated FGF-20 activity. Alternatively, the effectiveness of an

agent determined by a screening assay to decrease FGF-20 gene expression, protein levels, or downregulate FGF-20 activity, can be monitored in clinical trials of subjects exhibiting increased FGF-20 gene expression, protein levels, or upregulated FGF-20 activity. In such clinical trials, the expression or activity of an FGF-20 gene, and preferably, other genes that have been implicated in, for example, an FGF-20-associated disorder can be used as a "read out" or markers of the phenotype of a particular cell.

For example, and not by way of limitation, genes, including FGF-20, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) which modulates FGF-20 activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on FGF-20-associated disorders (*e.g.*, disorders characterized by deregulated cell proliferation and/or migration), for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of FGF-20 and other genes implicated in the FGF-20-associated disorder, respectively. The levels of gene expression (*e.g.*, a gene expression pattern) can be quantified by northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of FGF-20 or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during treatment of the individual with the agent.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) including the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of an FGF-20 protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the FGF-20 protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the FGF-20 protein, mRNA, or genomic DNA in the pre-administration sample with the FGF-20 protein, mRNA, or genomic DNA in the

post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of FGF-20 to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of FGF-20 to lower levels than detected, *i.e.* to decrease the effectiveness of the agent. According to such an embodiment, FGF-20 expression or activity may be used as an indicator of the effectiveness of an agent, even in the absence of an observable phenotypic response.

D. Methods of Treatment:

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant or unwanted FGF-20 expression or activity, *e.g.* a proliferative disorder, a differentiative disorder, a pain disorder. With regards to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. "Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (*e.g.*, a patient's "drug response phenotype", or "drug response genotype".) Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the FGF-20 molecules of the present invention or FGF-20 modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

1. Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant or unwanted FGF-20 expression or activity, by administering to the subject an FGF-20 or an agent which modulates FGF-20 expression or at least one FGF-20 activity. Subjects at risk for a disease which is

caused or contributed to by aberrant or unwanted FGF-20 expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the FGF-20 aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of FGF-20 aberrancy, for example, an FGF-20, FGF-20 agonist or FGF-20 antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

2. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating FGF-20 expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell with an FGF-20 or agent that modulates one or more of the activities of FGF-20 protein activity associated with the cell. An agent that modulates FGF-20 protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of an FGF-20 protein (*e.g.*, an FGF-20 substrate), an FGF-20 antibody, an FGF-20 agonist or antagonist, a peptidomimetic of an FGF-20 agonist or antagonist, or other small molecule. In one embodiment, the agent stimulates one or more FGF-20 activities. Examples of such stimulatory agents include active FGF-20 protein and a nucleic acid molecule encoding FGF-20 that has been introduced into the cell. In another embodiment, the agent inhibits one or more FGF-20 activities. Examples of such inhibitory agents include antisense FGF-20 nucleic acid molecules, anti-FGF-20 antibodies, and FGF-20 inhibitors. These modulatory methods can be performed *in vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant or unwanted expression or activity of an FGF-20 protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, upregulates or downregulates) FGF-20 expression or activity. In another embodiment, the method involves administering an FGF-20 protein or nucleic acid molecule as

therapy to compensate for reduced, aberrant, or unwanted FGF-20 expression or activity.

Stimulation of FGF-20 activity is desirable in situations in which FGF-20 is abnormally downregulated and/or in which increased FGF-20 activity is likely to have a beneficial effect. Likewise, inhibition of FGF-20 activity is desirable in situations in which FGF-20 is abnormally upregulated and/or in which decreased FGF-20 activity is likely to have a beneficial effect.

3. Pharmacogenomics

The FGF-20 molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on FGF-20 activity (*e.g.*, FGF-20 gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) FGF-20-associated disorders (*e.g.*, proliferative disorders) associated with aberrant or unwanted FGF-20 activity. In conjunction with such treatment, pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer an FGF-20 molecule or FGF-20 modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with an FGF-20 molecule or FGF-20 modulator.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, for example, Eichelbaum, M. *et al.* (1996) *Clin. Exp. Pharmacol. Physiol.* 23(10-11): 983-985 and Linder, M.W. *et al.* (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example,

glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (*e.g.*, a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants.) Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high resolution map can be generated from a combination of some ten-million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 1000 bases of DNA. A SNP may be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that may be common among such genetically similar individuals.

Alternatively, a method termed the "candidate gene approach", can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drugs target is known (*e.g.*, an FGF-20 protein of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation

as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Alternatively, a method termed the "gene expression profiling", can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (*e.g.*, an FGF-20 molecule or FGF-20 modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with an FGF-20 molecule or FGF-20 modulator, such as a modulator identified by one of the exemplary screening assays described herein.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application, as well as the Figures and the Sequence Listing, are incorporated herein by reference.

EXAMPLES

EXAMPLE 1: IDENTIFICATION AND CHARACTERIZATION OF FGF-20 cDNAs

In this example, the identification and characterization of the genes encoding monkey FGF-20 (clone jlkxb019e05) and human FGF-20 (clone fbh38777) is described.

Isolation of the FGF-20 cDNAs

The invention is based, at least in part, on the discovery of a monkey gene and a human gene encoding a novel protein, referred to herein as FGF-20. A clone was originally identified from a macaque dorsal root ganglion cDNA library using SEQUENCE EXPLORER™ software. The entire sequence of the monkey clone was determined and found to contain an open reading frame termed monkey "FGF-20."

The nucleotide sequence encoding the monkey FGF-20 protein is shown in Figure 1 and is set forth as SEQ ID NO:1. The protein encoded by this nucleic acid comprises about 177 amino acids and has the amino acid sequence shown in Figure 1 and set forth as SEQ ID NO:2. The coding region (open reading frame) of SEQ ID NO:1 is set forth as SEQ ID NO:3. Clone jlkxb019e05, comprising the partial coding region of monkey FGF-20 was deposited with the American Type Culture Collection (ATCC®), 10801 University Boulevard, Manassas, VA 20110-2209, on _____, and assigned Accession No. _____.

Further homology searching using a BLASTN 1.4.9 search, using a score of 100 and a word length of 12 (Altschul *et al.* (1990) *J. Mol. Biol.* 215:403), of the monkey FGF-20 sequence revealed that a human genomic fragment (GenBank Accession Number AC008012) comprised significant regional homology to monkey FGF-20. The exons in the human genomic sequence corresponding to the monkey FGF-20 cDNA were identified and assembled, and are referred to herein as human FGF-20.

The nucleotide sequence encoding the human FGF-20 protein, as identified within the Homo sapiens 12p13 BAC RPCI11-388F6 genomic fragment (Accession Number AC008012), is shown in Figure 9 and is set forth as SEQ ID NO:4. The protein encoded by this nucleic acid comprises about 178 amino acids and has the amino acid

sequence shown in Figure 9 and set forth as SEQ ID NO:5. The coding region (open reading frame) of SEQ ID NO:4 is set forth as SEQ ID NO:6.

A human fetal heart cDNA library was screened using a 394 bp probe corresponding to a fragment of human FGF-20, generated by PCR using the following primers derived from the human genomic sequence:

forward primer: GCCTTCCTGCCAGGCATGAACC (SEQ ID NO:10)
reverse primer: CTTTCCCATCCTCGGAACGTCAAGG (SEQ ID NO:11)

The library screening resulted in the identification of a positive clone (clone fbh38777) which was then sequenced. The human FGF-20 cDNA sequence is shown in Figure 10 and is set forth as SEQ ID NO:7. The protein encoded by this nucleic acid comprises about 178 amino acids and has the amino acid sequence shown in Figure 10 and set forth as SEQ ID NO:8. The coding region (open reading frame) of SEQ ID NO:7 is set forth as SEQ ID NO:9.

Analysis of the monkey FGF-20 Molecules

A BLASTP 1.4 search, using a score of 100 and a word length of 3 (Altschul et al. (1990) *J. Mol. Biol.* 215:403) of the protein sequence of monkey FGF-20 revealed that monkey FGF-20 is similar to the mouse FGF-15 protein (Accession Number AF007268) and the Human FGF-19 protein (Accession Number AB018122). The monkey FGF-20 protein is 38% identical to the mouse FGF-15 protein (Accession Number AF007268) over amino acid residues 9 to 68 and 42% identical to the human FGF-19 protein (Accession Number AB018122) over amino acid residues 9 to 69.

A BLASTN 1.4.9 search, using a score of 100 and a word length of 12 (Altschul *et al.* (1990) *J. Mol. Biol.* 215:403) of the nucleotide sequence of monkey FGF-20 revealed that monkey FGF-20 is similar to Mus musculus cDNA clone 619448 (Accession Number AA175629) and to Mycobacterium tuberculosis H37Rv complete genome; segment 126/162 (Accession Number Z74024). The monkey FGF-20 nucleic acid molecule is 74% identical to Mus musculus cDNA clone 619448 (Accession Number AA175629) over nucleotides 580 to 629. The monkey FGF-20 nucleic acid

molecule is 60% identical to Mycobacterium tuberculosis H37Rv complete genome; segment 126/162 (Accession Number Z74024) over nucleotides 302 to 397.

A search was performed against the HMM database resulting in the identification of a fibroblast growth factor domain in the amino acid sequence of monkey FGF-20 (SEQ ID NO:2) at about residues 1-55 of SEQ ID NO:2. The results of the search are set forth in Figure 3.

The monkey FGF-20 nucleic acid sequence was globally aligned with the Mus musculus mRNA (Accession Number AA175629) using the ALIGN program (version 2.0), using a PAM120 scoring matrix, a gap length penalty of 12 and a gap penalty of 4. The results showed a 32.2% identity (see Figure 4).

The monkey FGF-20 protein was globally aligned with the mouse fibroblast growth factor 15 (FGF-15) protein using the ALIGN program (version 2.0), using a PAM120 scoring matrix, a gap length penalty of 12 and a gap penalty of 4. The results showed a 14.7% identity (see Figure 5).

The monkey FGF-20 protein was globally aligned with the human fibroblast growth factor 19 (FGF-19) protein using the ALIGN program (version 2.0), using a PAM120 scoring matrix, a gap length penalty of 12 and a gap penalty of 4. The results showed a 17.4% identity (see Figure 6).

The monkey FGF-20 protein was also locally aligned with the mouse fibroblast growth factor 15 (FGF-15) protein using the LALIGN program (version 2.0u54), using a PAM120 scoring matrix, a gap length penalty of 12 and a gap penalty of 4. The results showed a 35.1% identity in a 74 amino acid overlap (see Figure 7).

Finally, the monkey FGF-20 protein was locally aligned with the human fibroblast growth factor 19 (FGF-19) protein using the LALIGN program (version 2.0u54), using a PAM120 scoring matrix, a gap length penalty of 12 and a gap penalty of 4. The results showed a 39.7% identity over a 78 amino acid overlap (see Figure 8).

Analysis of the human FGF-20 Molecules

A BLASTP 1.4 search, using a score of 50 and a word length of 3 (Altschul *et al.* (1990) *J. Mol. Biol.* 215:403) of the protein sequence of human FGF-20 revealed that human FGF-20 is similar to the human FGF-19 protein (Accession Number AB018122) and the mouse FGF-15 protein (Accession Number AF007268). The human FGF-20

protein is 40% identical to the human FGF-19 protein (Accession Number AB018122) over translated nucleotides 353 to 535 and 35% identical to the mouse FGF-15 protein (Accession Number AF007268) over translated nucleotides 353 to 532.

A BLASTN 1.4.9 search, using a score of 100 and a word length of 12 (Altschul *et al.* (1990) *J. Mol. Biol.* 215:403) of the nucleotide sequence of human FGF-20 revealed that the human FGF-20 nucleic acid molecule is 100% identical to residues 131413-131833 of Homo sapiens 12p13 BAC RPCI11-388F6 (Accession Number AC008012) over nucleotides 1-421. The human FGF-20 nucleic acid molecule is also 100% identical to residues 133644-135971 of Homo sapiens 12p13 BAC RPCI11-388F6 (Accession Number AC008012) over nucleotides 422-2749.

The human FGF-20 gene (SEQ ID NO:4) comprises sequences within nucleotide residues 131413-135971 of the Homo sapiens 12p13 BAC RPCI11-388F6 genomic fragment (Accession Number AC008012), with an intron present at about residues 131834-133643. Analysis of the Homo sapiens 12p13 BAC RPCI11-388F6 genomic sequence identified a consensus splice donor site sequence (GTGAGT) at nucleotide residues 131834-131839, a consensus splice acceptor site (TCCAG) at nucleotide residues 133639-133643, and a polyadenylation signal sequence (AATAAA) at residues 135950-135955.

Sequence analysis of human FGF-20 cDNA (SEQ ID NO:7)identified the same open reading frame (ORF) as the predicted ORF derived from the human genomic sequence ((SEQ ID NO:4); see Figure 12), and there are no differences in the nucleic acid sequence where the genomic and cDNA sequences overlap (see Figure 11).

A search was performed against the HMM database resulting in the identification of a fibroblast growth factor domain in the amino acid sequence of human FGF-20 (SEQ ID NO:5 or 8) at about residues 2-56 of SEQ ID NO:5 or 8. The results of the search are set forth in Figure 14.

The human FGF-20 protein is predicted to have at least one cAMP and cGMP dependent protein kinase phosphorylation site, at about amino acid residues 102-105 of SEQ ID NO:5 or 8.

The human FGF-20 protein is predicted to have at least one protein kinase C phosphorylation site, at about amino acid residues 17-19, 86-88 and 112-114 of SEQ ID NO:5 or 8.

The human FGF-20 protein is predicted to have at least one casein kinase II phosphorylation site, at about amino acid residues 107-110, 112-115, 139-142 and 166-169 of SEQ ID NO:5 or 8.

The human FGF-20 protein was globally aligned with the human fibroblast growth factor-19 (FGF-19) protein using the GAP program in the GCG software package, using a Blosum 62 matrix and a gap weight of 12 and a length weight of 4. The results showed a 29.6% identity (see Figure 15). Also indicated was the presence of a conserved cysteine residue in human FGF-20, at about amino acid 40 of SEQ ID NO:5 or 8, which corresponds to cysteine 120 of human FGF-19.

The human FGF-20 protein was globally aligned with the mouse fibroblast growth factor-15 (FGF-15) protein using the GAP program in the GCG software package, using a Blosum 62 matrix and a gap weight of 12 and a length weight of 4. The results showed a 22.3% identity (see Figure 16). Also indicated was the presence of a conserved cysteine residue in human FGF-20, at about amino acid 40 of SEQ ID NO:5 or 8, which corresponds to cysteine 127 of mouse FGF-15.

The human FGF-20 nucleic acid sequence was globally aligned with the monkey fibroblast growth factor-20 (FGF-20) nucleic acid sequence using the GAP program in the GCG software package, using a nwsgapdna matrix and a gap weight of 12 and a length weight of 4. The results showed a 94.5% identity between the two sequences (see Figure 17).

Finally, the human FGF-20 protein was globally aligned with the monkey fibroblast growth factor-20 (FGF-20) protein using the GAP program in the GCG software package, using a Blosum 62 matrix and a gap weight of 12 and a length weight of 4. The results showed a 93.8% identity (see Figure 18).

Tissue Distribution of FGF-20 mRNA

This example describes the tissue distribution of FGF-20 mRNA, as was determined by Polymerase Chain Reaction (PCR) on cDNA libraries using oligonucleotide primers specific to the human FGF-20 sequence.

The human FGF-20 gene is expressed in human fetal heart and heart tissue from a subject suffering from congestive heart failure, fetal liver, trigeminal ganglion, bone marrow, as well as in human tumors (*e.g.*, colon to liver metastases and erythroleukemia cells).

EXAMPLE 2: EXPRESSION OF RECOMBINANT FGF-20 PROTEIN IN BACTERIAL CELLS

In this example, FGF-20 is expressed as a recombinant glutathione-S-transferase (GST) fusion polypeptide in *E. coli* and the fusion polypeptide is isolated and characterized. Specifically, FGF-20 is fused to GST and this fusion polypeptide is expressed in *E. coli*, *e.g.*, strain PEB199. Expression of the GST-FGF-20 fusion protein in PEB199 is induced with IPTG. The recombinant fusion polypeptide is purified from crude bacterial lysates of the induced PEB199 strain by affinity chromatography on glutathione beads. Using polyacrylamide gel electrophoretic analysis of the polypeptide purified from the bacterial lysates, the molecular weight of the resultant fusion polypeptide is determined.

EXAMPLE 3: EXPRESSION OF RECOMBINANT FGF-20 PROTEIN IN COS CELLS

To express the FGF-20 gene in COS cells, the pcDNA/Amp vector by Invitrogen Corporation (San Diego, CA) is used. This vector contains an SV40 origin of replication, an ampicillin resistance gene, an *E. coli* replication origin, a CMV promoter followed by a polylinker region, and an SV40 intron and polyadenylation site. A DNA fragment encoding the entire FGF-20 protein and an HA tag (Wilson *et al.* (1984) *Cell* 37:767) or a FLAG tag fused in-frame to its 3' end of the fragment is cloned into the polylinker region of the vector, thereby placing the expression of the recombinant protein under the control of the CMV promoter.

To construct the plasmid, the FGF-20 DNA sequence is amplified by PCR using two primers. The 5' primer contains the restriction site of interest followed by approximately twenty nucleotides of the FGF-20 coding sequence starting from the

initiation codon; the 3' end sequence contains complementary sequences to the other restriction site of interest, a translation stop codon, the HA tag or FLAG tag and the last 20 nucleotides of the FGF-20 coding sequence. The PCR amplified fragment and the pCDNA/Amp vector are digested with the appropriate restriction enzymes and the vector is dephosphorylated using the CIAP enzyme (New England Biolabs, Beverly, MA). Preferably the two restriction sites chosen are different so that the FGF-20 gene is inserted in the correct orientation. The ligation mixture is transformed into *E. coli* cells (strains HB101, DH5α, SURE, available from Stratagene Cloning Systems, La Jolla, CA, can be used), the transformed culture is plated on ampicillin media plates, and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment.

COS cells are subsequently transfected with the FGF-20-pcDNA/Amp plasmid DNA using the calcium phosphate or calcium chloride co-precipitation methods, DEAE-dextran-mediated transfection, lipofection, or electroporation. Other suitable methods for transfecting host cells can be found in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989. The expression of the FGF-20 polypeptide is detected by radiolabelling (^{35}S-methionine or ^{35}S-cysteine available from NEN, Boston, MA, can be used) and immunoprecipitation (Harlow, E. and Lane, D. *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988) using an HA specific monoclonal antibody. Briefly, the cells are labelled for 8 hours with ^{35}S-methionine (or ^{35}S-cysteine). The culture media are then collected and the cells are lysed using detergents (RIPA buffer, 150 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% DOC, 50 mM Tris, pH 7.5). Both the cell lysate and the culture media are precipitated with an HA specific monoclonal antibody. Precipitated polypeptides are then analyzed by SDS-PAGE.

Alternatively, DNA containing the FGF-20 coding sequence is cloned directly into the polylinker of the pCDNA/Amp vector using the appropriate restriction sites. The resulting plasmid is transfected into COS cells in the manner described above, and the expression of the FGF-20 polypeptide is detected by radiolabelling and immunoprecipitation using an FGF-20 specific monoclonal antibody.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

What is claimed:

1. An isolated nucleic acid molecule selected from the group consisting of:

(a) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:1, 4, or 7; and

(b) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:3, 6, or 9.

2. An isolated nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 2, 5, or 8.

3. An isolated nucleic acid molecule comprising the nucleotide sequence contained in the plasmid deposited with ATCC® as Accession Number _____.

4. An isolated nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 2, 5, or 8.

5. An isolated nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence which is at least 60% identical to the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9, or a complement thereof;

b) a nucleic acid molecule comprising a fragment of at least 107 nucleotides of a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9, or a complement thereof;

c) a nucleic acid molecule which encodes a polypeptide comprising an amino acid sequence at least about 50% identical to the amino acid sequence of SEQ ID NO:2, 5, or 8; and

d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 5, or 8, wherein the fragment comprises at least 15 contiguous amino acid residues of the amino acid sequence of SEQ ID NO: 2, 5, or 8.

6. An isolated nucleic acid molecule which hybridizes to the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 under stringent conditions.

7. An isolated nucleic acid molecule comprising a nucleotide sequence which is complementary to the nucleotide sequence of the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5.

8. An isolated nucleic acid molecule comprising the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5, and a nucleotide sequence encoding a heterologous polypeptide.

9. A vector comprising the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5.

10. The vector of claim 9, which is an expression vector.

11. A host cell transfected with the expression vector of claim 10.

12. A method of producing a polypeptide comprising culturing the host cell of claim 11 in an appropriate culture medium to, thereby, produce the polypeptide.

13. An isolated polypeptide selected from the group consisting of:

a) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 5, or 8, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO: 2, 5, or 8;

b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2 or 5, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule consisting of SEQ ID NO:1, 3, 4, 6, 7, or 9 under stringent conditions;

c) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 60 % identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9;

d) a polypeptide comprising an amino acid sequence which is at least 50% identical to the amino acid sequence of SEQ ID NO: 2, 5, or 8.

14. The isolated polypeptide of claim 13 comprising the amino acid sequence of SEQ ID NO:2, 5, or 8.

15. The polypeptide of claim 13, further comprising heterologous amino acid sequences.

16. An antibody which selectively binds to a polypeptide of claim 13.

17. A method for detecting the presence of a polypeptide of claim 13 in a sample comprising:

a) contacting the sample with a compound which selectively binds to the polypeptide; and

b) determining whether the compound binds to the polypeptide in the sample to thereby detect the presence of a polypeptide of claim 13 in the sample.

18. The method of claim 17, wherein the compound which binds to the polypeptide is an antibody.

19. A kit comprising a compound which selectively binds to a polypeptide of claim 13 and instructions for use.

20. A method for detecting the presence of a nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 in a sample comprising:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and

b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample to thereby detect the presence of a nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 in the sample.

21. The method of claim 20, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

22. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 and instructions for use.

23. A method for identifying a compound which binds to a polypeptide of claim 13 comprising:

a) contacting the polypeptide, or a cell expressing the polypeptide with a test compound; and

b) determining whether the polypeptide binds to the test compound.

24. The method of claim 23, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

a) detection of binding by direct detection of test compound/polypeptide binding;

b) detection of binding using a competition binding assay; and

c) detection of binding using an assay for FGF-20 activity.

25. A method for modulating the activity of a polypeptide of claim 13 comprising contacting the polypeptide or a cell expressing the polypeptide with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

26. A method for identifying a compound which modulates the activity of a polypeptide of claim 13 comprising:

a) contacting a polypeptide of claim 13 with a test compound; and

b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

```
V  R  S  E  D  A  G  F  V  V  I  T  G  V  M  S  R  R  Y          19
C GTC CGA TCA GAG GAT GCT GGC TTT GTG GTG ATT ACA GGT GTG ATG AGC AGA AGA TAC    58

 L  C  M  D  F  R  G  N  I  F  G  S  H  Y  F  N  P  E  N  C    39
CTC TGC ATG GAT TTC AGA GGC AAC ATT TTT GGA TCA CAC TAT TTC AAC CCG GAG AAC TGC  118

 R  F  R  H  W  T  L  E  N  G  Y  D  V  Y  H  S  P  Q  H  H    59
AGG TTC CGA CAC TGG ACG CTG GAG AAC GGC TAC GAC GTC TAC CAC TCT CCT CAG CAT CAC  178

 F  L  V  S  L  G  R  A  K  R  A  F  L  P  G  M  N  P  P  P    79
TTT CTG GTC AGT CTG GGC CGG GCG AAG AGG GCC TTC CTG CCA GGC ATG AAC CCA CCC CCC  238

 Y  S  Q  F  L  S  R  R  N  E  I  P  L  I  H  F  N  T  P  R    99
TAC TCC CAG TTC CTG TCC CGG AGG AAC GAG ATC CCC CTC ATC CAC TTC AAT ACC CCC AGA  298

 P  R  R  H  T  R  S  A  E  D  E  S  E  R  D  P  L  N  V  L   119
CCA CGG CGG CAC ACC CGG AGC GCC GAG GAC GAG TCG GAG CGG GAC CCC CTG AAC GTG CTG  358

 K  P  R  A  R  M  T  P  A  P  A  S  C  S  Q  E  L  P  S  A   139
AAG CCC CGG GCC CGG ATG ACC CCG GCC CCG GCC TCC TGC TCA CAG GAG CTC CCG AGC GCC  418

 E  D  N  S  P  V  A  S  D  P  L  G  V  V  R  G  G  R  V  N   159
GAG GAC AAC AGC CCG GTG GCC AGC GAC CCG TTA GGG GTG GTC AGG GGC GGT CGG GTG AAC  478

 T  H  A  G  G  T  G  P  E  A  C  R  P  F  P  K  F  I  *      178
ACG CAC GCT GGG GGA ACG GGC CCG GAA GCC TGC CGC CCC TTC CCC AAG TTC ATC TAG      535

GGTGGCTGGAAGGGCACCCTCTTTAACCCATCCCTCAGCATAGCAAGCTCTTCCAAGGACCAAGCTCCTTGACGTTCCG  614

AGGATGGGAAAGGTGACAGGGGCAATGTATGGAATTGCTGCTTCTCTGGGGTCCCTTCCACAGGAGGTCCTTGTGAGAA  693

TCAACCTTTAGGCCCAAGTCATGGGGTTTCAACANCTTTCTTCACTTCAACATAGAACAACCTTTTCCGAATAGGAAAC  772

CCCGACAGGTAAACTAGNAATTTTCCCCTTTAT                                              805
```

FIGURE 1

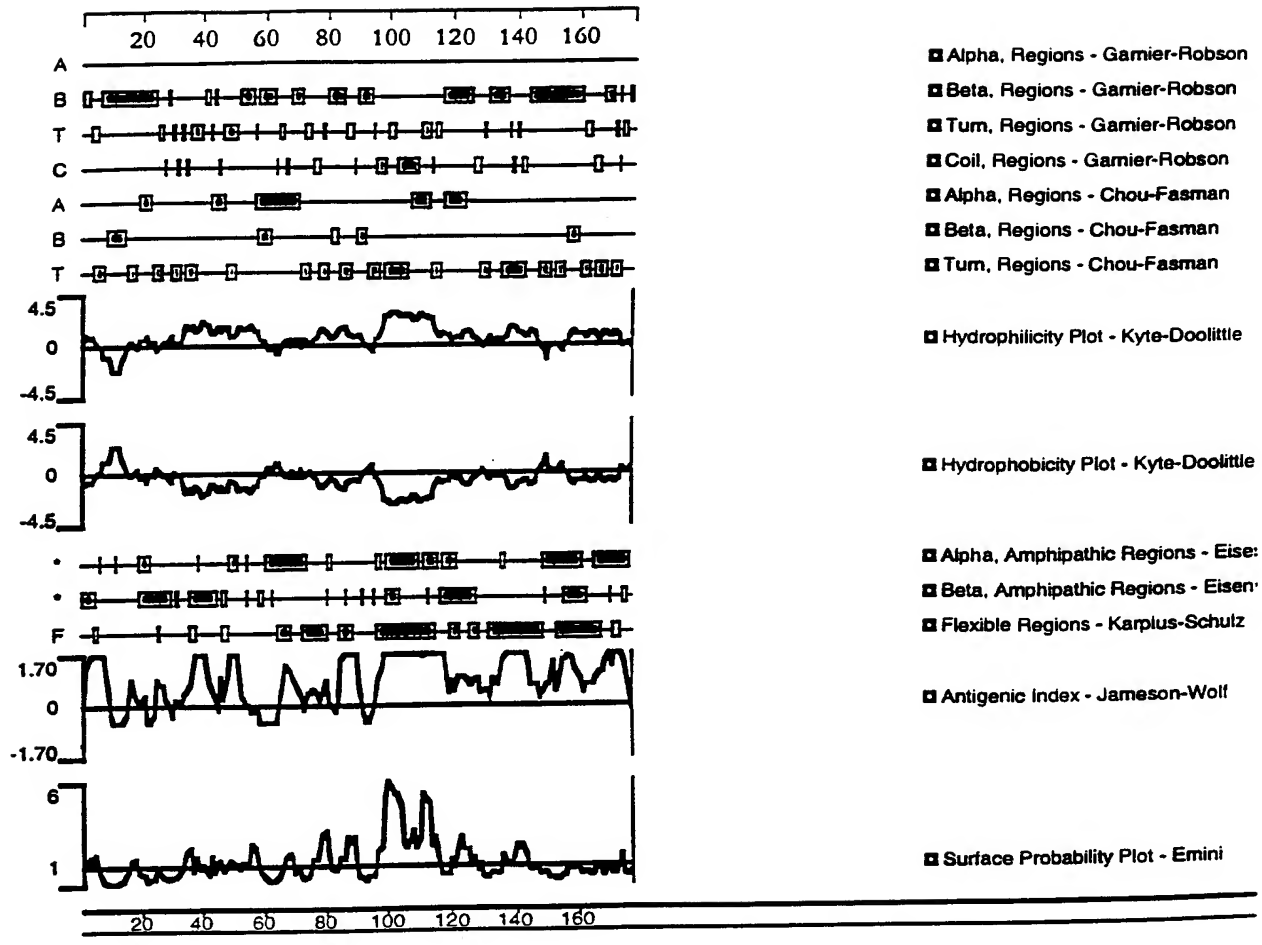
20 40 60 80 100 120 140 160
A
B
T
C
A
B
T
4.5
0
-4.5
4.5
0
-4.5
F
1.70
0
-1.70
6
1
20 40 60 80 100 120 140 160
Alpha, Regions - Garnier-Robson
Beta, Regions - Garnier-Robson
Turn, Regions - Garnier-Robson
Coil, Regions - Garnier-Robson
Alpha, Regions - Chou-Fasman
Beta, Regions - Chou-Fasman
Turn, Regions - Chou-Fasman
Hydrophilicity Plot - Kyte-Doolittle
Hydrophobicity Plot - Kyte-Doolittle
Alpha, Amphipathic Regions - Eise
Beta, Amphipathic Regions - Eisen
Flexible Regions - Karplus-Schulz
Antigenic Index - Jameson-Wolf
Surface Probability Plot - Emini

FIGURE 2

```
5.1 Protein Family / Domain Matches, HMMer version 2

Searching for complete domains
hmmpfam - search a single seq against HMM database
HMMER 2.1.1 (Dec 1998)
Copyright (C) 1992-1998 Washington University School of Medicine
HMMER is freely distributed under the GNU General Public License (GPL).
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
HMM file:                 /prod/ddm/seqanal/PFAM/pfam3.4/Pfam
Sequence file:            /tmp/orfanal.3598.aa
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 Query:  jlkxbl9e5

Scores for sequence family classification (score includes all domains):
Model    Description                                        Score    E-value
N
-------- -----------                                        -----    -------
---
FGF      PF00167 Fibroblast growth factor                    24.3      5e-06
1

Parsed for domains:
Model    Domain  seq-f seq-t    hmm-f hmm-t       score  E-value
-------- ------- ----- -----    ----- -----       -----  -------
FGF        1/1       1    55 [.    40    94 ..    24.3    5e-06

Alignments of top-scoring domains:
FGF: domain 1 of 1, from 1 to 55: score 24.3, E = 5e-06
                   *->isaverGiVsIrGveSglYLAMnkkGkLYASkkGltee.CvFrErle
                      +++++ G V+I Gv S++YL+M+ +G   +S   ++ e+C Fr   +
   jlkxbl9e5     1    VRSEDAGFVVITGVMSRRYLCMDFRGNIFGSHY-FNPEnCRFRHWTL 46

                   eNnYNTYaS<-*
                   eN+Y  Y S
   jlkxbl9e5    47 ENGYDVYHS    55
```

FIGURE 3

```
ALIGN calculates a global alignment of two sequences
 version 2.0uPlease cite: Myers and Miller, CABIOS (1989)
> j1kxbl9e5pepnuc                                          805 aa vs.
> Genbank AA175629 - ms96g05.r1 Soares mouse 3NbM          447 aa
scoring matrix: pam120.mat, gap penalties: -12/-4
32.2% identity;     Global alignment score: -421

                10        20        30        40        50        60
inputs CGTCCGATCAGAGGATGCTGGCTTTGTGGTGATTACAGGTGTGATGAGCAGAAGATACCT
        .: :           :.:.. :::::    :.:::  .:: : :.:.  :::::::
       GATTC----------TACAA-CTTTG----GTTTA--AGTTTTAAGTT-AGAAGAT----
                       10              20          30

                70        80        90        100       110       120
inputs CTGCATGGATTTCAGAGGCAACATTTTTGGATCACACTATTTCAACCCGGAGAACTGCAG
        :: .:::::.: ..:::::.:  :::::::...:   .:::::.::
       -TG-TTGGATATTTAAGGCTA--TTTTTAATT---TCTATTACA----------------
        40          50          60          70

               130       140       150       160       170       180
inputs GTTCCGACACTGGACGCTGGAGAACGGCTACGACGTCTACCACTCTCCTCAGCATCACTT
        :.:.:      :: .:.::::  :.: .: :. :.: . . :::.::       ::
       -----GTCTC------CTTAAAAAC--CAAAAAGGAATGCATTAATCCACA-------TT
              80                90        100       110

               190       200       210       220       230       240
inputs TCTGGTCAGTCTGGGCCGGGCGAAGAGGGCCTTCCTGCCAGGCATGAACCCACCCCCCTA
       ::                          :::::::  ::.. .:::.          ..
       TC--------------------------CCTTCCT--CAAAAGTGTA----------AT
                                   120       130

               250       260       270       280       290       300
inputs CTCCCAGTTCCTGTCCCGGAGGAACGAGATCCCCCTCATCCACTTCAATACCCCCAGACC
        ::: .: :::: ...:::. .::..    : .  :.::      :::: ..:
       GTCCTTGGTCCTTGG--AAGGGATTAAGGATATTATAGGACGCT------CCCCAGAATT
        140       150       160       170             180

               310       320       330       340       350       360
inputs ACGGCGGCACACCCGGAGCGCCGAGGACGAGTCGGAGCGGGACCCCCTGAACGTGCTGAA
       .:.:: ::.:: :.:    : :::.       :.::                  :::...
       GCAGCTGCTCATACAG----CTGAGA-------GAAG----------------GCTATT
      190       200       210                                220

               370       380       390       400       410       420
inputs GCCCCGGGCCCGGATGACCCCGGCCCCGGCCTCCTGCTCACAGGAGCTCCCGAGCGCCGA
       .          .:::    : ::: .    :.::::::.  .            .
       A----------TTGA-----GTCCCTA-----TACTCACTTTT-------------T
                           230       240

               430       440       450       460       470       480
```

FIGURE 4

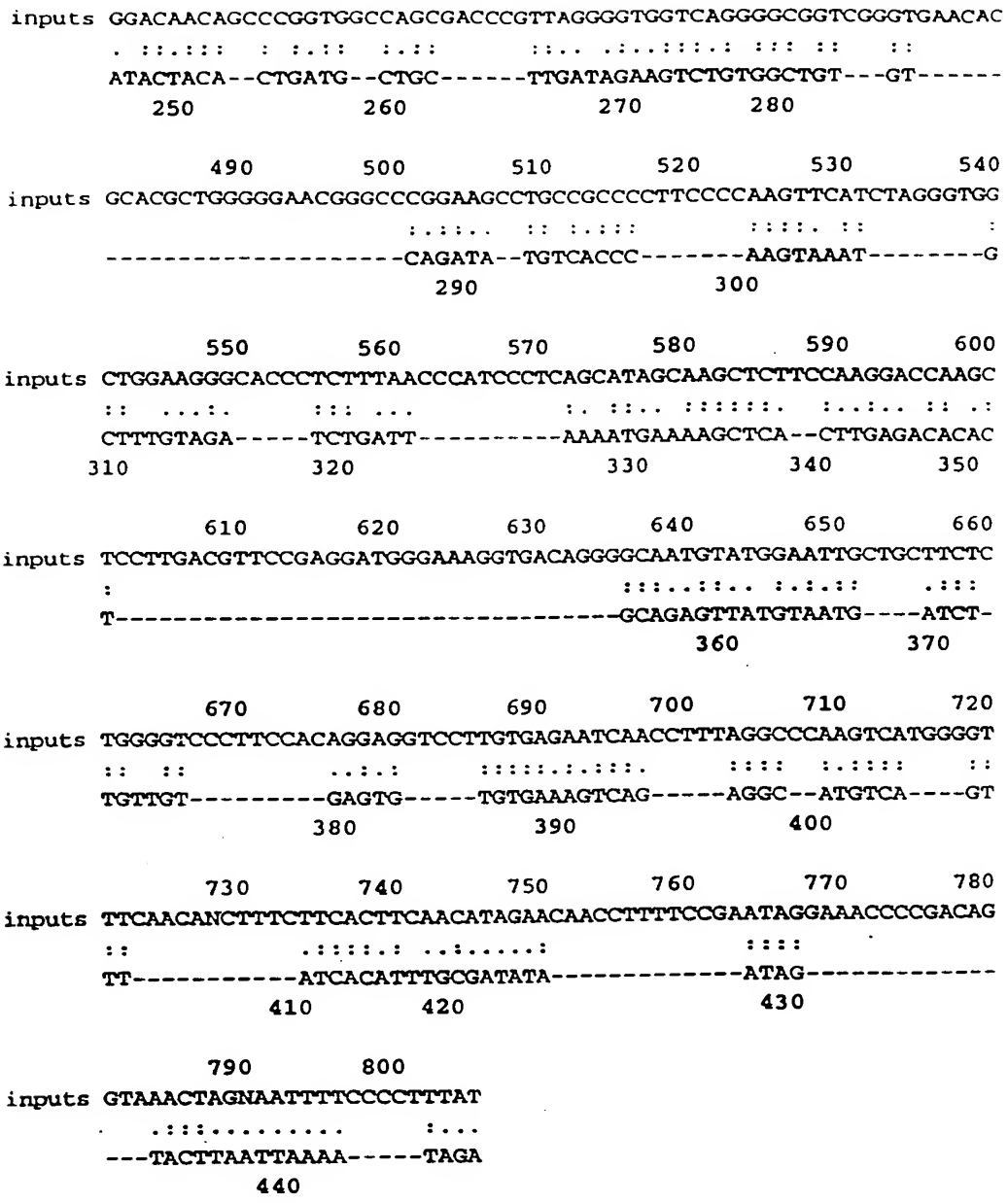

FIGURE 4 (cont'd)

```
ALIGN calculates a global alignment of two sequences
 version 2.0uPlease cite: Myers and Miller, CABIOS (1989)
> jlkxbl9e5pep                                          177 aa vs.
> Genbank O35622 - (mouse FGF-15)                        218 aa
scoring matrix: pam120.mat, gap penalties: -12/-4
14.7% identity;  Global alignment score: -248

                10         20         30         40              50
inputs VRSEDAGFVVITG-VMSRRYLCMDFRGNIFGSHYFNPENCRFRH-WT----LENGYDVYH
       .    .: .:  . :..  : .. :     :.  . :.  :   :     :.  :..
       MARKWNGRAVARALVLATLWLAVSGRPLAQQSQSVSDEDPLFLYGWGKITRLQYLYSAGP
                10         20         30         40         50         60

            60                  70                  80          90
inputs SPQHHFL-----VSLG----RAKRAFLP---------GMNPPPYSQFLSRRNE--I-PLI
          . ::       :..     .. :.:          ...  .  ..:   ..  :  ::
       YVSNCFLRIRSDGSVDCEEDQNERNLLEFRAVALKTIAIKDVSSVRYLCMSADGKIYGLI
                70         80         90        100        110        120

                100        110             120        130        140
inputs HFN----TPRPRRHTRSAED-ESERDPLNVL----KPRARM-TPAPASCSQELPSA--ED
       ...    : :   .  .  ..  : .. :...    :::...  . :.. . .  .  :
       RYSEEDCTFREEMDCLGYNQYRSMKHHLHIIFIQAKPREQLQDQKPSNFIPVFHRSFFET
               130        140        150        160        170        180

              150        160        170
inputs NSPVASDPLGV-VRGGRVNTHAGGTGPEACRPFPKF-I
       .... :  ... . .. ..      . .      : :
       GDQLRSKMFSLPLESDSMDPFRMVEDVDHLVKSPSFQK
               190        200        210
```

FIGURE 5

```
ALIGN calculates a global alignment of two sequences
 version 2.0uPlease cite: Myers and Miller, CABIOS (1989)
> jlkxb19e5pep                                          177 aa vs.
> FGF-19                                                216 aa
scoring matrix: pam120.mat, gap penalties: -12/-4
17.4% identity;  Global alignment score: -220

                 10          20          30          40
inputs VRSE-DAGFV-VITGV-MSRRYLCMDFRGNIFGSHYFNPENCRFRH--------------
       .::   .  : ....:. ..     ..: ..    ::   .  :.::
       MRSGCVVVHVWILAGLWLAVAGRPLAFSDAGPHVHYGWGDPIRLRHLYTSGPHGLSSCFL
                10          20          30          40          50          60

            50          60          70          80
inputs WTLENGY-DVYHSPQHHFLVSL-GRAKRAF-LPGMNPPPY--------SQFLSR------
       .. ...:  :  ...  : :. . . : :.  . :... :          : : .
       RIRADGVVDCARGQSAHSLLEIKAVALRTVAIKGVHSVRYLCMGADGKMQGLLQYSEEDC
                70          80          90          100         110         120

            90         100         110         120         130         140
inputs --RNEIPLIHFNTPRPRRHTRSAEDESERDPLNVLKPRARMTPAPASCSQELPSAEDN--
         .::    .:. :. .:   ..   :... .. : :. .  .       . ..: .
       AFEEEIRPDGYNVYRSEKHRLPVSLSSAKQR-QLYKNRGFLPLSHFLPMLPMVPEEPEDL
                130         140         150         160         170

             150         160         170
inputs -SPVASDPLGV-VRGGRVNTHAGGTGPEACRPFPKF-I
       . ..:: ..  .  . . ..  .  :: :: :. : :
       RGHLESDMFSSPLETDSMDPFGLVTGLEAVRS-PSFEK
     180         190         200         210
```

FIGURE 6

```
LALIGN finds the best local alignments between two sequences
 version 2.0u54 July 1996
Please cite:
 X. Huang and W. Miller (1991) Adv. Appl. Math. 12:373-381

 Comparison of:
(A) inputs/nb945222.tmp > jlkxbl9e5pep
- 177 aa
(B) inputs/nb22224.tmp > Genbank O35622 - Fibroblast Growth Factor-15
- 218 aa
 using matrix file: pam120.mat, gap penalties: -12/-4

  35.1% identity in 74 aa overlap; score:   94

      10         20         30         40         50         60
      VVITGVMSRRYLCMDFRGNIFGSHYFNPENCRFRHWTLE-NGYDVYHSPQHHFLVSLGRA
      ..: .: : ::::::.  :.:.:    .. :.: ::    ..  ::. :.: .::. . . .:
      IAIKDVSSVRYLCMSADGKIYGLIRYSEEDCTFRE-EMDCLGYNQYRSMKHHLHIIFIQA
       100        110        120        130        140        150

        70         80
      K-RAFLPGMNPPPY
      : :. :.. .:. .
      KPREQLQDQKPSNF
        160
```

FIGURE 7

```
 LALIGN finds the best local alignments between two sequences
 version 2.0u54 July 1996
Please cite:
 X. Huang and W. Miller (1991) Adv. Appl. Math. 12:373-381

 Comparison of:
(A) inputs/nb588513.tmp > jlkxb19e5pep                                  - 177 aa
(B) inputs/nb704015.tmp > FGF-19                                        - 216 aa
 using matrix file: pam120.mat, gap penalties: -12/-4

  39.7% identity in 78 aa overlap; score:  108

       10        20        30        40        50        60
       VVITGVMSRRYLCMDFRGNIFGSHYFNPENCRFRHWTLENGYDVYHSPQHHFLVSLGRAK
       :.: :: : :::::.  :.. :   .. :.: :   . .::.::.: .:.. :::. ::
       VAIKGVHSVRYLCMGADGKMQGLLQYSEEDCAFEEEIRPDGYNVYRSEKHRLPVSLSSAK
      90        100       110       120       130       140

       70        80
       -RAFLPGMNPPPYSQFLS
        : .  . .  : :.::.
       QRQLYKNRGFLPLSHFLP
     150       160

----------

  66.7% identity in 6 aa overlap; score:   27

       70
       KRAFLP
       .:.:::
       NRGFLP
         160

----------

  42.9% identity in 14 aa overlap; score:   26

             90
       QFLSRRNEIPLIHF
       :.   :. .:: ::
       QLYKNRGFLPLSHF
            160

----------

  62.5% identity in 8 aa overlap; score:   25

      150
       DPLGVVRG
```

FIGURE 8

```
gggaaaataa agacggtcca ggtagagaga gagaaacatg tgttcagcac aggtagaaga 60

attccaggag ctcagagtgc cccatacagg caacaagatg aagcaggagg tgaatgactg 120

tatgtgtgtt gggggcaaga gaggatgtca gaagaaacgc tgaatatgca gaaatgaggc 180

tgaatttaag agtgctgaag ttatcaccac ccttaaaatc aatccaggga ggtttcatga 240

aggtaggttt tcaggaggtg cttgaaggtg ggaattggat ggcaatgagt ctttgccctg 300

cctgtttttc tccataggtg ccctg atg atc aga tca gag gat gct ggc ttt   352
                           Met Ile Arg Ser Glu Asp Ala Gly Phe
                             1               5

gtg gtg att aca ggt gtg atg agc aga aga tac ctc tgc atg gat ttc   400
Val Val Ile Thr Gly Val Met Ser Arg Arg Tyr Leu Cys Met Asp Phe
 10                  15                  20                  25

aga ggc aac att ttt gga tca cac tat ttc gac ccg gag aac tgc agg   448
Arg Gly Asn Ile Phe Gly Ser His Tyr Phe Asp Pro Glu Asn Cys Arg
                 30                  35                  40

ttc caa cac cag acg ctg gaa aac ggg tac gac gtc tac cac tct cct   496
Phe Gln His Gln Thr Leu Glu Asn Gly Tyr Asp Val Tyr His Ser Pro
             45                  50                  55

cag tat cac ttc ctg gtc agt ctg ggc cgg gcg aag aga gcc ttc ctg   544
Gln Tyr His Phe Leu Val Ser Leu Gly Arg Ala Lys Arg Ala Phe Leu
         60                  65                  70

cca ggc atg aac cca ccc ccg tac tcc cag ttc ctg tcc cgg agg aac   592
Pro Gly Met Asn Pro Pro Pro Tyr Ser Gln Phe Leu Ser Arg Arg Asn
     75                  80                  85

gag atc ccc cta att cac ttc aac acc ccc ata cca cgg cgg cac acc   640
Glu Ile Pro Leu Ile His Phe Asn Thr Pro Ile Pro Arg Arg His Thr
 90                  95                 100                 105

cgg agc gcc gag gac gac tcg gag cgg gac ccc ctg aac gtg ctg aag   688
Arg Ser Ala Glu Asp Asp Ser Glu Arg Asp Pro Leu Asn Val Leu Lys
                110                 115                 120

ccc cgg gcc cgg atg acc ccg gcc ccg gcc tcc tgt tca cag gag ctc   736
Pro Arg Ala Arg Met Thr Pro Ala Pro Ala Ser Cys Ser Gln Glu Leu
            125                 130                 135

ccg agc gcc gag gac aac agc ccg atg gcc agt gac cca tta ggg gtg   784
Pro Ser Ala Glu Asp Asn Ser Pro Met Ala Ser Asp Pro Leu Gly Val
        140                 145                 150

gtc agg ggc ggt cga gtg aac acg cac gct ggg gga acg ggc ccg gaa   832
Val Arg Gly Gly Arg Val Asn Thr His Ala Gly Gly Thr Gly Pro Glu
    155                 160                 165

ggc tgc cgc ccc ttc gcc aag ttc atc tag ggtcgctgga agggcaccct     882
Gly Cys Arg Pro Phe Ala Lys Phe Ile
170                 175
```

FIGURE 9

ctttaaccca tccctcagca aacgcagctc ttcccaagga ccaggtccct tgacgttccg 942

aggatgggaa aggtgacagg ggcatgtatg gaatttgctg cttctctggg gtcccttcca 1002

caggaggtcc tgtgagaacc aacctttgag gcccaagtca tggggtttca ccgccttcct 1062

cactccatat agaacacctt tcccaatagg aaaccccaac aggtaaacta gaaatttccc 1122

cttcatgaag gtagagagaa ggggtctctc ccaacatatt tctcttcctt gtgcctctcc 1182

tctttatcac ttttaagcat aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aagcagtggg 1242

ttcctgagct caagactttg aaggtgtagg gaagaggaaa tcggagatcc cagaagcttc 1302

tccactgccc tatgcattta tgttagatgc cccgatccca ctggcatttg agtgtgcaaa 1362

ccttgacatt aacagctgaa tggggcaagt tgatgaaaac actactttca agccttcgtt 1422

cttccttgag catctctggg gaagagctgt caaaagactg gtggtaggct ggtgaaaact 1482

tgacagctag acttgatgct tgctgaaatg aggcaggaat cataatagaa aactcagcct 1542

ccctacaggg tgagcacctt ctgtctcgct gtctccctct gtgcagccac agccagaggg 1602

cccagaatgg ccccactctg ttcccaagca gttcatgata cagcctcacc ttttggcccc 1662

atctctggtt tttgaaaatt tggtctaagg aataaatagc ttttacactg gctcacgaaa 1722

atctgccctg ctagaatttg cttttcaaaa tggaaataaa ttccaactct cctaagaggc 1782

atttaattaa ggctctactt ccaggttgag taggaatcca ttctgaacaa actacaaaaa 1842

tgtgactggg aaggggggctt tgagagactg ggactgctct gggttaggtt ttctgtggac 1902

tgaaaaatcg tgtccttttc tctaaatgaa gtggcatcaa ggactcaggg ggaaagaaat 1962

caggggacat gttatagaag ttatgaaaag acaaccacat ggtcaggctc ttgtctgtgg 2022

tctctagggc tctgcagcag cagtggctct tcgattagtt aaaactctcc taggctgaca 2082

catctgggtc tcaatcccct tggaaattct tggtgcatta aatgaagcct taccccatta 2142

ctgcggttct tcctgtaagg gggctccatt ttcctccctc tctttaaatg accacctaaa 2202

ggacagtata ttaacaagca aagtcgattc aacaacagct tcttcccagt cacttttttt 2262

tttctcactg ccatcacata ctaaccttat actttgatct attcttttg gttatgagag 2322

aaatgttggg caactgtttt tacctgatgg ttttaagctg aacttgaagg actggttcct 2382

attctgaaac agtaaaacta tgtataatag tatatagcca tgcatggcaa atattttaat 2442

atttctgttt tcatttcctg ttggaaatat tatcctgcat aatagctatt ggaggctcct 2502

cagtgaaaga tcccaaaagg attttggtgg aaaactagtt gtaatctcac aaactcaaca 2562

FIGURE 9
CONT.

ctaccatcag gggttttctt tatggcaaag ccaaaatagc tcctacaatt tcttatatcc 2622

ctcgtcatgt ggcagtattt atttatttat ttggaagttt gcctatcctt ctatatttat 2682

agatatttat aaaaatgtaa ccccttttc ctttcttctg tttaaaataa aaataaaatt 2742

tatctca 2749

FIGURE 9
CONT.

```
ccacgcgtcc ggtggggaag aaatctcgct gaattatcac gcatgttaca ccagtatatg 60

atctaattgt gcctttgcca caaaacagta atttaaagcc attatcaatt acttaagagg 120

taggtcgtgt gaatgggttt caggcccttg tcggagacta gtttttgaga ggggacactg 180

aaagtccatg aggggctgca cctggagagg tcaccaccaa gtgagaaaat gacaaagaac 240

caacccaaga agagccaaga agaaaattcc atccgtcact tatattgatt caacataaac 300

agttataccc tctgctccta agcagctcac tctaaggaac gcactggata ggtaaactca 360

gctaaagcaa gttaaatgga atacatgctg taatagaggt gaaggcattg tcctgaggag 420

ctgagaagga agaacaactg attttgaatg gaaagatgag gaaagtcttc atagagatgg 480

tgacgcctga gcctggtctt gaagagtgag tgacttcaat aagtagagaa ggaagaggga 540

gatcaactct actaccattc tgtacacata ctgggtgttg actgatgtat tagacaatta 600

cacagacatc caggaggaga atcagactct atggcaagct ggatccttga aagacatctc 660

agcatagatt taaaaatcac aaagtagaag gcatggaaga atgtgactat caccacaaac 720

attcaaaggt attagtaagg caaaagggaa aataaagacg gtccaggtag agagagagaa 780

acatgtgttc agcacaggta gaagaattcc aggagctcag agtgccccat acaggcaaca 840

agatgaagca ggaggtgaat gactgtatgt gtgttggggg caagagagga tgtcagaaga 900

aacgctgaat atgcagaaat gaggctgaat ttaagagtgc tgaagttatc accaccctta 960

aaatcaatcc agggaggttt catgaaggta ggttttcagg aggtgcttga aggtgggaat 1020

tggatggcaa tgagtctttg ccctgcctgt ttttctccat aggtgccctg atg atc    1076
                                                    Met Ile
                                                      1

aga tca gag gat gct ggc ttt gtg gtg att aca ggt gtg atg agc aga    1124
Arg Ser Glu Asp Ala Gly Phe Val Val Ile Thr Gly Val Met Ser Arg
          5                  10                  15

aga tac ctc tgc atg gat ttc aga ggc aac att ttt gga tca cac tat    1172
Arg Tyr Leu Cys Met Asp Phe Arg Gly Asn Ile Phe Gly Ser His Tyr
     20                  25                  30

ttc gac ccg gag aac tgc agg ttc caa cac cag acg ctg gaa aac ggg    1220
Phe Asp Pro Glu Asn Cys Arg Phe Gln His Gln Thr Leu Glu Asn Gly
 35                  40                  45                  50
```

FIGURE 10

```
tac gac gtc tac cac tct cct cag tat cac ttc ctg gtc agt ctg ggc   1268
Tyr Asp Val Tyr His Ser Pro Gln Tyr His Phe Leu Val Ser Leu Gly
                55                  60                  65

cgg gcg aag aga gcc ttc ctg cca ggc atg aac cca ccc ccg tac tcc   1316
Arg Ala Lys Arg Ala Phe Leu Pro Gly Met Asn Pro Pro Pro Tyr Ser
            70                  75                  80

cag ttc ctg tcc cgg agg aac gag atc ccc cta att cac ttc aac acc   1364
Gln Phe Leu Ser Arg Arg Asn Glu Ile Pro Leu Ile His Phe Asn Thr
        85                  90                  95

ccc ata cca cgg cgg cac acc cgg agc gcc gag gac gac tcg gag cgg   1412
Pro Ile Pro Arg Arg His Thr Arg Ser Ala Glu Asp Asp Ser Glu Arg
    100                 105                 110

gac ccc ctg aac gtg ctg aag ccc cgg gcc cgg atg acc ccg gcc ccg   1460
Asp Pro Leu Asn Val Leu Lys Pro Arg Ala Arg Met Thr Pro Ala Pro
115                 120                 125                 130

gcc tcc tgt tca cag gag ctc ccg agc gcc gag gac aac agc ccg atg   1508
Ala Ser Cys Ser Gln Glu Leu Pro Ser Ala Glu Asp Asn Ser Pro Met
                135                 140                 145

gcc agt gac cca tta ggg gtg gtc agg ggc ggt cga gtg aac acg cac   1556
Ala Ser Asp Pro Leu Gly Val Val Arg Gly Gly Arg Val Asn Thr His
            150                 155                 160

gct ggg gga acg ggc ccg gaa ggc tgc cgc ccc ttc gcc aag ttc atc   1604
Ala Gly Gly Thr Gly Pro Glu Gly Cys Arg Pro Phe Ala Lys Phe Ile
        165                 170                 175

tagggtcgct ggaagggcac cctctttaac ccatccctca gcaaacgcag ctcttcccaa 1664

ggaccaggtc ccttgacgtt ccgaggatgg gaaaggtgac aggggcatgt atggaatttg 1724

ctgcttctct ggggtccctt ccacaggagg tcctgtgaga accaaccttt gaggcccaag 1784

tcatggggtt tcaccgcctt cctcactcca tatagaacac ctttcccaat aggaaacccc 1844

aacaggtaaa ctagaaattt ccccttcatg aaggtagaga gaaggggtct ctcccaacat 1904

atttctcttc cttgtgcctc tcctctttat cacttttaag cataaaaaaa aaaaaaaaaa 1964

aaaaaaaaa                                                         1973
```

FIGURE 10, CONT.

CLUSTAL W (1.74) multiple sequence alignment

```
hFGF20.cDNA          CCACGCGTCCGGTGGGGAAGAAATCTCGCTGAATTATCACGCATGTTACACCAGTATATG
hFGF20.genomic       ------------------------------------------------------------

hFGF20.cDNA          ATCTAATTGTGCCTTTGCCACAAAACAGTAATTTAAAGCCATTATCAATTACTTAAGAGG
hFGF20.genomic       ------------------------------------------------------------

hFGF20.cDNA          TAGGTCGTGTGAATGGGTTTCAGGCCCTTGTCGGAGACTAGTTTTTGAGAGGGGACACTG
hFGF20.genomic       ------------------------------------------------------------

hFGF20.cDNA          AAAGTCCATGAGGGGCTGCACCTGGAGAGGTCACCACCAAGTGAGAAAATGACAAAGAAC
hFGF20.genomic       ------------------------------------------------------------

hFGF20.cDNA          CAACCCAAGAAGAGCCAAGAAGAAAATTCCATCCGTCACTTATATTGATTCAACATAAAC
hFGF20.genomic       ------------------------------------------------------------

hFGF20.cDNA          AGTTATACCCTCTGCTCCTAAGCAGCTCACTCTAAGGAACGCACTGGATAGGTAAACTCA
hFGF20.genomic       ------------------------------------------------------------

hFGF20.cDNA          GCTAAAGCAAGTTAAATGGAATACATGCTGTAATAGAGGTGAAGGCATTGTCCTGAGGAG
hFGF20.genomic       ------------------------------------------------------------

hFGF20.cDNA          CTGAGAAGGAAGAACAACTGATTTTGAATGGAAAGATGAGGAAAGTCTTCATAGAGATGG
hFGF20.genomic       ------------------------------------------------------------

hFGF20.cDNA          TGACGCCTGAGCCTGGTCTTGAAGAGTGAGTGACTTCAATAAGTAGAGAAGGAAGAGGGA
hFGF20.genomic       ------------------------------------------------------------

hFGF20.cDNA          GATCAACTCTACTACCATTCTGTACACATACTGGGTGTTGACTGATGTATTAGACAATTA
hFGF20.genomic       ------------------------------------------------------------

hFGF20.cDNA          CACAGACATCCAGGAGGAGAATCAGACTCTATGGCAAGCTGGATCCTTGAAAGACATCTC
hFGF20.genomic       ------------------------------------------------------------

hFGF20.cDNA          AGCATAGATTTAAAAATCACAAAGTAGAAGGCATGGAAGAATGTGACTATCACCACAAAC
hFGF20.genomic       ------------------------------------------------------------

hFGF20.cDNA          ATTCAAAGGTATTAGTAAGGCAAAAGGGAAAATAAAGACGGTCCAGGTAGAGAGAGAGAA
hFGF20.genomic       -------------------------GGGAAAATAAAGACGGTCCAGGTAGAGAGAGAGAA
                                              **********************************

hFGF20.cDNA          ACATGTGTTCAGCACAGGTAGAAGAATTCCAGGAGCTCAGAGTGCCCCATACAGGCAACA
hFGF20.genomic       ACATGTGTTCAGCACAGGTAGAAGAATTCCAGGAGCTCAGAGTGCCCCATACAGGCAACA
                     ************************************************************

hFGF20.cDNA          AGATGAAGCAGGAGGTGAATGACTGTATGTGTGTTGGGGGCAAGAGAGGATGTCAGAAGA
```

FIGURE 11

```
hFGF20.genomic      AGATG..AGCAGGAGGTGAATGACTGTATGTGTGTTGC..GGCAAGAGAGGATGTCAGAAGA
                    ************************************************************

hFGF20.cDNA         AACGCTGAATATGCAGAAATGAGGCTGAATTTAAGAGTGCTGAAGTTATCACCACCCTTA
hFGF20.genomic      AACGCTGAATATGCAGAAATGAGGCTGAATTTAAGAGTGCTGAAGTTATCACCACCCTTA
                    ************************************************************

hFGF20.cDNA         AAATCAATCCAGGGAGGTTTCATGAAGGTAGGTTTTCAGGAGGTGCTTGAAGGTGGGAAT
hFGF20.genomic      AAATCAATCCAGGGAGGTTTCATGAAGGTAGGTTTTCAGGAGGTGCTTGAAGGTGGGAAT
                    ************************************************************

hFGF20.cDNA         TGGATGGCAATGAGTCTTTGCCCTGCCTGTTTTTCTCCATAGGTGCCCTGATGATCAGAT
hFGF20.genomic      TGGATGGCAATGAGTCTTTGCCCTGCCTGTTTTTCTCCATAGGTGCCCTGATGATCAGAT
                    ************************************************************

hFGF20.cDNA         CAGAGGATGCTGGCTTTGTGGTGATTACAGGTGTGATGAGCAGAAGATACCTCTGCATGG
hFGF20.genomic      CAGAGGATGCTGGCTTTGTGGTGATTACAGGTGTGATGAGCAGAAGATACCTCTGCATGG
                    ************************************************************

hFGF20.cDNA         ATTTCAGAGGCAACATTTTTGGATCACACTATTTCGACCCGGAGAACTGCAGGTTCCAAC
hFGF20.genomic      ATTTCAGAGGCAACATTTTTGGATCACACTATTTCGACCCGGAGAACTGCAGGTTCCAAC
                    ************************************************************

hFGF20.cDNA         ACCAGACGCTGGAAAACGGGTACGACGTCTACCACTCTCCTCAGTATCACTTCCTGGTCA
hFGF20.genomic      ACCAGACGCTGGAAAACGGGTACGACGTCTACCACTCTCCTCAGTATCACTTCCTGGTCA
                    ************************************************************

hFGF20.cDNA         GTCTGGGCCGGGCGAAGAGAGCCTTCCTGCCAGGCATGAACCCACCCCCGTACTCCCAGT
hFGF20.genomic      GTCTGGGCCGGGCGAAGAGAGCCTTCCTGCCAGGCATGAACCCACCCCCGTACTCCCAGT
                    ************************************************************

hFGF20.cDNA         TCCTGTCCCGGAGGAACGAGATCCCCCTAATTCACTTCAACACCCCCATACCACGGCGGC
hFGF20.genomic      TCCTGTCCCGGAGGAACGAGATCCCCCTAATTCACTTCAACACCCCCATACCACGGCGGC
                    ************************************************************

hFGF20.cDNA         ACACCCGGAGCGCCGAGGACGACTCGGAGCGGGACCCCCTGAACGTGCTGAAGCCCCGGG
hFGF20.genomic      ACACCCGGAGCGCCGAGGACGACTCGGAGCGGGACCCCCTGAACGTGCTGAAGCCCCGGG
                    ************************************************************

hFGF20.cDNA         CCCGGATGACCCCGGCCCCGGCCTCCTGTTCACAGGAGCTCCCGAGCGCCGAGGACAACA
hFGF20.genomic      CCCGGATGACCCCGGCCCCGGCCTCCTGTTCACAGGAGCTCCCGAGCGCCGAGGACAACA
                    ************************************************************

hFGF20.cDNA         GCCCGATGGCCAGTGACCCATTAGGGGTGGTCAGGGGCGGTCGAGTGAACACGCACGCTG
hFGF20.genomic      GCCCGATGGCCAGTGACCCATTAGGGGTGGTCAGGGGCGGTCGAGTGAACACGCACGCTG
                    ************************************************************

hFGF20.cDNA         GGGGAACGGGCCCGGAAGGCTGCCGCCCCTTCGCCAAGTTCATCTAGGGTCGCTGGAAGG
hFGF20.genomic      GGGGAACGGGCCCGGAAGGCTGCCGCCCCTTCGCCAAGTTCATCTAGGGTCGCTGGAAGG
                    ************************************************************

hFGF20.cDNA         GCACCCTCTTTAACCCATCCCTCAGCAAACGCAGCTCTTCCCAAGGACCAGGTCCCTTGA
hFGF20.genomic      GCACCCTCTTTAACCCATCCCTCAGCAAACGCAGCTCTTCCCAAGGACCAGGTCCCTTGA
                    ************************************************************

hFGF20.cDNA         CGTTCCGAGGATGGGAAAGGTGACAGGGGCATGTATGGAATTTGCTGCTTCTCTGGGGTC
hFGF20.genomic      CGTTCCGAGGATGGGAAAGGTGACAGGGGCATGTATGGAATTTGCTGCTTCTCTGGGGTC
                    ************************************************************

hFGF20.cDNA         CCTTCCACAGGAGGTCCTGTGAGAACCAACCTTTGAGGCCCAAGTCATGGGGTTTCACCG
```

FIGURE 11

CONT.

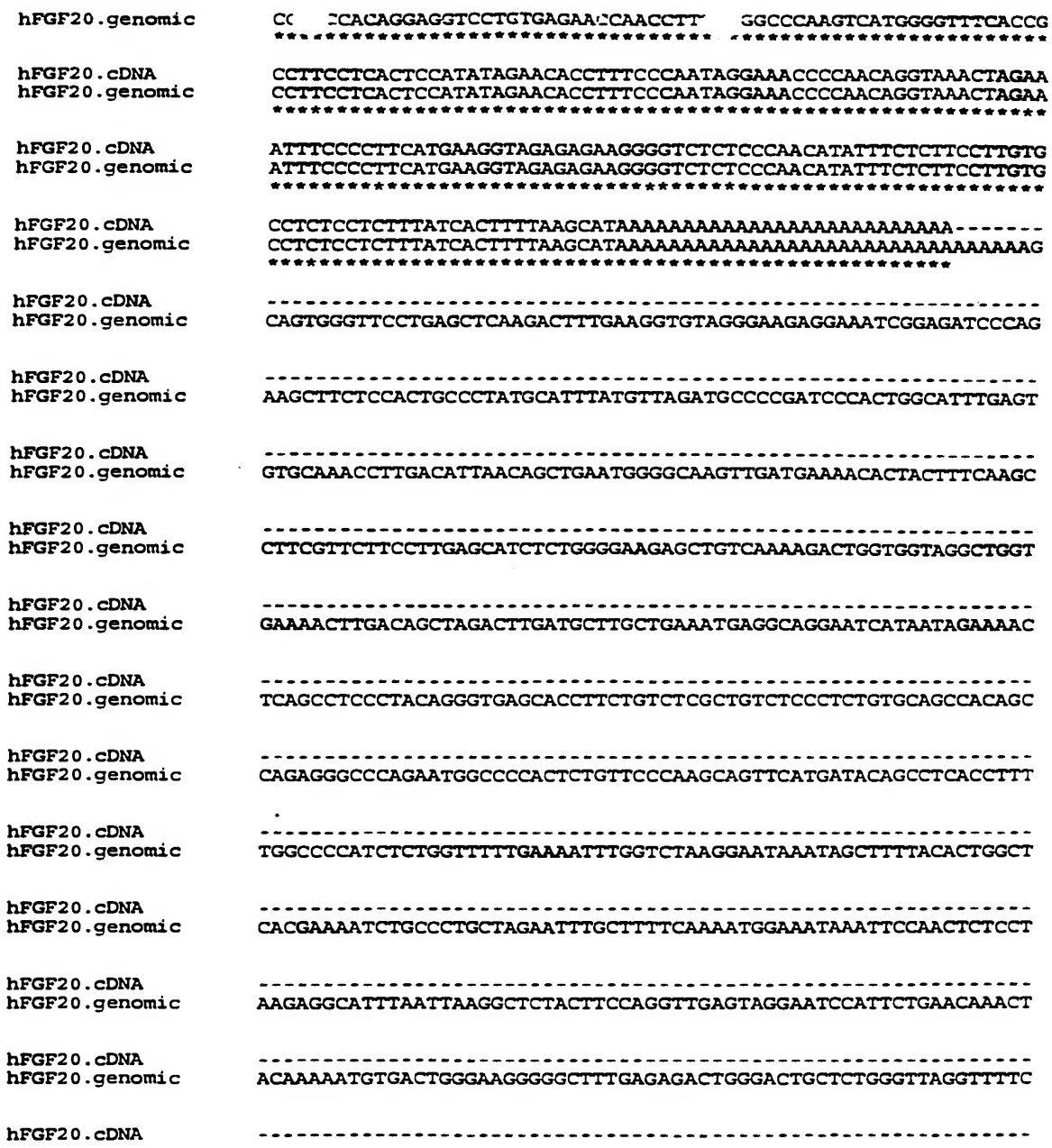

FIGURE 11

CONT.

```
hFGF20.genomic      TGTGGACTGAAAAATCGTGTCCTTTTCTCTAAATGAAGTGGCATCAAGGACTCAGGGGGA

hFGF20.cDNA         ------------------------------------------------------------
hFGF20.genomic      AAGAAATCAGGGGACATGTTATAGAAGTTATGAAAAGACAACCACATGGTCAGGCTCTTG

hFGF20.cDNA         ------------------------------------------------------------
hFGF20.genomic      TCTGTGGTCTCTAGGGCTCTGCAGCAGCAGTGGCTCTTCGATTAGTTAAAACTCTCCTAG

hFGF20.cDNA         ------------------------------------------------------------
hFGF20.genomic      GCTGACACATCTGGGTCTCAATCCCCTTGGAAATTCTTGGTGCATTAAATGAAGCCTTAC

hFGF20.cDNA         ------------------------------------------------------------
hFGF20.genomic      CCCATTACTGCGGTTCTTCCTGTAAGGGGGCTCCATTTTCCTCCCTCTCTTTAAATGACC

hFGF20.cDNA         ------------------------------------------------------------
hFGF20.genomic      ACCTAAAGGACAGTATATTAACAAGCAAAGTCGATTCAACAACAGCTTCTTCCCAGTCAC

hFGF20.cDNA         ------------------------------------------------------------
hFGF20.genomic      TTTTTTTTTTCTCACTGCCATCACATACTAACCTTATACTTTGATCTATTCTTTTTGGTT

hFGF20.cDNA         ------------------------------------------------------------
hFGF20.genomic      ATGAGAGAAATGTTGGGCAACTGTTTTTACCTGATGGTTTTAAGCTGAACTTGAAGGACT

hFGF20.cDNA         ------------------------------------------------------------
hFGF20.genomic      GGTTCCTATTCTGAAACAGTAAAACTATGTATAATAGTATATAGCCATGCATGGCAAATA

hFGF20.cDNA         ------------------------------------------------------------
hFGF20.genomic      TTTTAATATTTCTGTTTTCATTTCCTGTTGGAAATATTATCCTGCATAATAGCTATTGGA

hFGF20.cDNA         ------------------------------------------------------------
hFGF20.genomic      GGCTCCTCAGTGAAAGATCCCAAAAGGATTTTGGTGGAAAACTAGTTGTAATCTCACAAA

hFGF20.cDNA         ------------------------------------------------------------
hFGF20.genomic      CTCAACACTACCATCAGGGGTTTTCTTTATGGCAAAGCCAAAATAGCTCCTACAATTTCT

hFGF20.cDNA         ------------------------------------------------------------
hFGF20.genomic      TATATCCCTCGTCATGTGGCAGTATTTATTTATTTATTTGGAAGTTTGCCTATCCTTCTA

hFGF20.cDNA         ------------------------------------------------------------
hFGF20.genomic      TATTTATAGATATTTATAAAAATGTAACCCCTTTTTCCTTTCTTCTGTTTAAAATAAAAA

hFGF20.cDNA         --------------
hFGF20.genomic      TAAAATTTATCTCA
```

FIGURE 11

CONT.

```
CLUSTAL W (1.74) multiple sequence alignment

hFGF20_genomic_     MIRSEDAGFVVITGVMSRRYLCMDFRGNIFGSHYFDPENCRFQHQTLENGYDVYHSPQY
hFGF20_cDNA_        MIRSEDAGFVVITGVMSRRYLCMDFRGNIFGSHYFDPENCRFQHQTLENGYDVYHSPQY
                    ************************************************************

hFGF20_genomic_     FLVSLGRAKRAFLPGMNPPPYSQFLSRRNEIPLIHFNTPIPRRHTRSAEDDSERDPLNV
hFGF20_cDNA_        FLVSLGRAKRAFLPGMNPPPYSQFLSRRNEIPLIHFNTPIPRRHTRSAEDDSERDPLNV
                    ************************************************************

hFGF20_genomic_     KPRARMTPAPASCSQELPSAEDNSPMASDPLGVVRGGRVNTHAGGTGPEGCRPFAKFI
hFGF20_cDNA_        KPRARMTPAPASCSQELPSAEDNSPMASDPLGVVRGGRVNTHAGGTGPEGCRPFAKFI
                    ************************************************************
```

FIGURE 12

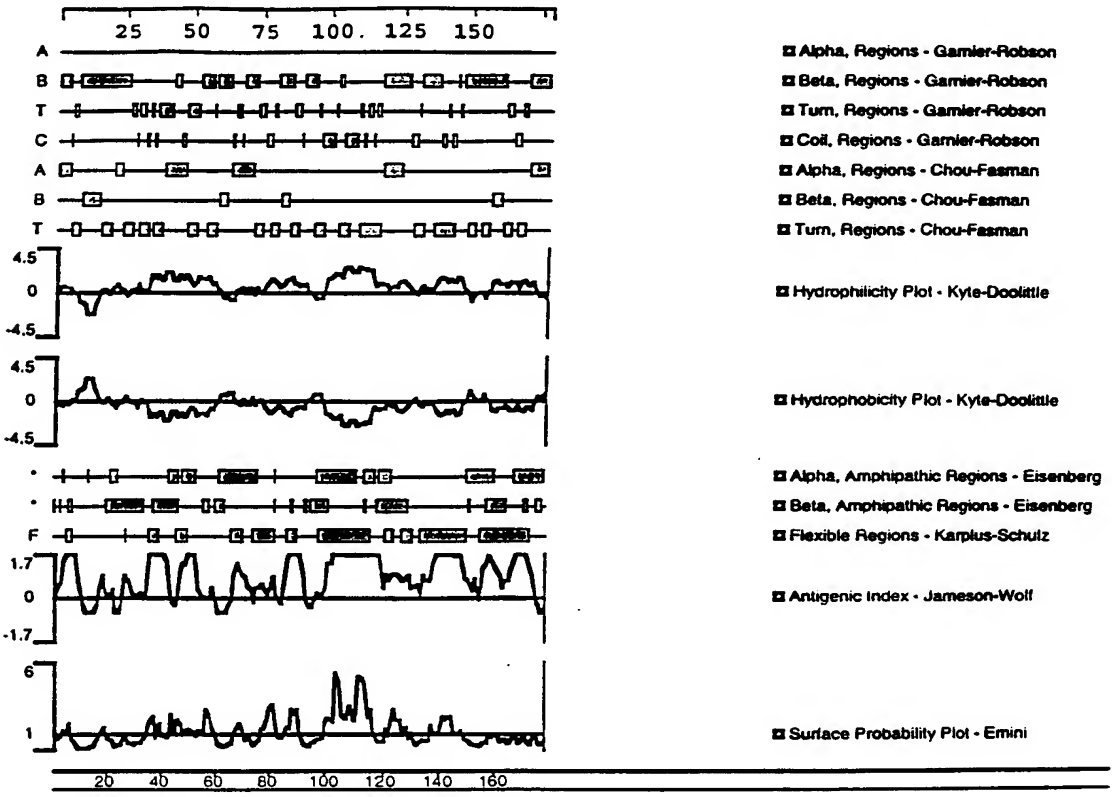
25 50 75 100 125 150
A
B
T
C
A
B
T
4.5
0
-4.5
4.5
0
-4.5
F
1.7
0
-1.7
6
1
20 40 60 80 100 120 140 160
Alpha, Regions - Garnier-Robson
Beta, Regions - Garnier-Robson
Turn, Regions - Garnier-Robson
Coil, Regions - Garnier-Robson
Alpha, Regions - Chou-Fasman
Beta, Regions - Chou-Fasman
Turn, Regions - Chou-Fasman
Hydrophilicity Plot - Kyte-Doolittle
Hydrophobicity Plot - Kyte-Doolittle
Alpha, Amphipathic Regions - Eisenberg
Beta, Amphipathic Regions - Eisenberg
Flexible Regions - Karplus-Schulz
Antigenic Index - Jameson-Wolf
Surface Probability Plot - Emini

FIGURE 13

Protein Family / Domain Matches, HMMer version 2

```
Searching for complete domains
hmmpfam - search a single seq against HMM database
HMMER 2.1.1 (Dec 1998)
Copyright (C) 1992-1998 Washington University School of Medicine
HMMER is freely distributed under the GNU General Public License (GPL).
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
HMM file:                 /prod/ddm/seqanal/PFAM/pfam4.2/Pfam
Sequence file:            /usr/ns-home/docs/seqanal/orfanal/oa-script.7819.seq
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
  Query:  human

Scores for sequence family classification (score includes all domains):
Model    Description                                        Score    E-value  N
-------- -----------                                        -----    ------- ---
FGF      Fibroblast growth factor                            21.6      3e-05   1

Parsed for domains:
Model    Domain  seq-f seq-t    hmm-f hmm-t       score  E-value
-------- ------- ----- -----    ----- -----       -----  -------
FGF        1/1       2    56 ..     40    94 ..    21.6    3e-05

Alignments of top-scoring domains:
FGF: domain 1 of 1, from 2 to 56: score 21.6, E = 3e-05
                  *->issaverGiVsIrGveSglYLAMnkkGkLYASkkGltee.CvFrErle
                     i++++ G V+I Gv S++YL+M+ +G    +S    +  e+C F  + +
       human      2    IRSEDAGFVVITGVMSRRYLCMDFRGNIFGSHY-FDPEnCRFQHQTL 47

                    eNnYNTYaS<-*
                    eN+Y  Y S
       human     48 ENGYDVYHS     56

//
```

ProDom Matches

```
ProDom entry 549        Match length 136
Keywords: GROWTH FACTOR FIBROBLAST PRECURSOR MITOGEN SIGNAL HEPARIN-BINDING GLYCOPROTEIN PROTEIN VASCULARIZATION

Expect 4.0e-05  Score 105        Bits 45.3        Identical 0.38  Conserved 0.57
query  2 IRSEDAGFVVITGVMSRRYLCMDFRGNIFGSHYFDPENCRFQHQTLENGYDVYHSPQY
         I + + G V I GV S  YLCM+ +G ++G       E+C F+ +  EN Y+ Y S +Y
sbjct 40 ISAVEMGVVSIKGVESGHYLCMNKKGKLYGKSKFTEDCVFRERIEENNYNTYASRKY
```

FIGURE 14

```
 GAP of: humanfgf19.pep  check: 9162  from: 1  to: 216

humanFGF19 4514718 in GenPept

 to: humanfgf20.pep  check: 851  from: 1  to: 178

humanFGF20 (analysis only) - Import - complete

 Symbol comparison table: /ddm_local/gcg/gcg_9.1/gcgcore/data/rundata/blosum62.c
 CompCheck: 6430

        Gap Weight:     12       Average Match:  2.912
     Length Weight:      4    Average Mismatch: -2.003

           Quality:     83              Length:     259
             Ratio:  0.466               Gaps:       3
 Percent Similarity: 37.037    Percent Identity: 29.630

       Match display thresholds for the alignment(s):
                   | = IDENTITY
                   : =    2
                   . =    1

 humanfgf19.pep x humanfgf20.pep


      51 GPHGLSSCFLRIRADGVVDCARGQSAHSLLEIKAVALRTVAIKGVHSVRY 100
                                        |:.     | | || | ||
       1 ..........................MIRSEDAGFVVITGVMSRRY 20

     101 LCMGADGKMQGLLQYSEEDCAFEEEIRPDGYNVYRSEKHRLPVSLSSAKQ 150
         |||   | . |   :  |.| |: :   .||.|| | .:   |||  ||
      21 LCMDFRGNIFGSHYFDPENCRFQHQTLENGYDVYHSPQYHFLVSLGRAK. 69

     151 RQLYKNRGFLPLSHFLPM...LPMVPEEPEDLRGHLESDMFSSPLE.TDS 196
         |      | | ||     :|::      | |  |     |  :  .
      70 RAFLPGMNPPPYSQFLSRRNEIPLIHFNTPIPRRHTRSAEDDSERDPLNV 119

     197 MDPFGLVTGLEAVRSPSFEK.............................. 216
         : |   .|   |  |
     120 LKPRARMTPAPASCSQELPSAEDNSPMASDPLGVVRGGRVNTHAGGTGPE 169
```

FIGURE 15

```
 GAP of: humanfgf.  pep  check: 851  from: 1  to:  ?8

humanFGF20 (analysis only) - Import - complete

 to: mousefgf15.pep  check: 6968  from: 1  to: 218

mouseFGF15 Genbank O35622 - (untitled)

 Symbol comparison table: /ddm_local/gcg/gcg_9.1/gcgcore/data/rundata/blosum62.c
 CompCheck: 6430

        Gap Weight:     12       Average Match:  2.912
     Length Weight:      4   Average Mismatch: -2.003

           Quality:     58             Length:     266
             Ratio:  0.326               Gaps:       2
 Percent Similarity: 30.769   Percent Identity: 22.308

       Match display thresholds for the alignment(s):
                  | = IDENTITY
                  : =   2
                  . =   1

humanfgf20.pep x mousefgf15.pep


                .         .         .         .         .
  1 .....................................MIRSEDAGFVVIT 13
                                         |.    : |
 51 RLQYLYSAGPYVSNCFLRIRSDGSVDCEEDQNERNLLEFRAVALKTIAIK 100
                .         .         .         .         .
 14 GVMSRRYLCMDFRGNIFGSHYFDPENCRFQHQTLENGYDVYHSPQYHFLV 63
     | | |||||   | |:|   :  |.| |. :    ||. | | .:|  :
101 DVSSVRYLCMSADGKIYGLIRYSEEDCTFREEMDCLGYNQYRSMKHHLHI 150
                .         .         .         .         .
 64 SLGRAK.RAFLPGMNPPPYSQFLSRRNEIPLIHFNTPIPRRHTRSAEDDS 112
    .|| |  |    |  :    |          :  :       :
151 IFIQAKPREQLQDQKPSNFIPVFHR....SFFETGDQLRSKMFSLPLESD 196
                .         .         .         .         .
113 ERDPLNVLKPRARMTPAPASCSQELPSAEDNSPMASDPLGVVRGGRVNTH 162
      ||  ...    :  .|.
197 SMDPFRMVEDVDHLVKSPSFQK.......................... 218
                                  .
                                  .
                                  .
```

FIGURE 16

```
GAP of: KBAa000fk  check: 6563  from: 1  to: 2733

               (analysis only) - Import - complete

to: LBAa000fk  check: 6230  from: 1  to: 2574

monkey FGF20.seq (analysis only) - Import - complete

Symbol comparison table: /ddm_local/gcg/gcg_9.1/gcgcore/data/rundata/nwsgapdna.cmp
CompCheck: 8760

        Gap Weight:      12      Average Match: 10.000
     Length Weight:       4   Average Mismatch:  0.000

           Quality:  22236              Length:   2892
             Ratio:  8.639                Gaps:      5
Percent Similarity: 94.493   Percent Identity: 94.493

       Match display thresholds for the alignment(s):
                  | = IDENTITY
                  : =    5
                  . =    1

KBAa000fk x LBAa000fk
```

```
                .         .         .         .         .
301 GCCTGTTTTTCTCCATAGGTGCCCTGATGATCAGATCAGAGGATGCTGGC 350   hFGF-20
                          |  | ||       || ||||||||||||||||||
  1 ..................GTCGACCCACGCGTCCGATCAGAGGATGCTGGC 33    monkeyFGF-20
                .         .         .         .         .
351 TTTGTGGTGATTACAGGTGTGATGAGCAGAAGATACCTCTGCATGGATTT 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 34 TTTGTGGTGATTACAGGTGTGATGAGCAGAAGATACCTCTGCATGGATTT 83
                .         .         .         .         .
401 CAGAGGCAACATTTTTGGATCACACTATTTCGACCCGGAGAACTGCAGGT 450
    |||||||||||||||||||||||||||||| ||||||||||||||||||
 84 CAGAGGCAACATTTTTGGATCACACTATTTCAACCCGGAGAACTGCAGGT 133
                .         .         .         .         .
451 TCCAACACCAGACGCTGGAAAACGGGTACGACGTCTACCACTCTCCTCAG 500
    ||| ||||  |||||||||| ||||| ||||||||||||||||||||||||
134 TCCGACACTGGACGCTGGAGAACGGCTACGACGTCTACCACTCTCCTCAG 183
                .         .         .         .         .
501 TATCACTTCCTGGTCAGTCTGGGCCGGGCGAAGAGAGCCTTCCTGCCAGG 550
    ||||||| ||||||||||||||||||||||||| ||||||||||||||||
184 CATCACTTTCTGGTCAGTCTGGGCCGGGCGAAGAGGGCCTTCCTGCCAGG 233
                .         .         .         .         .
551 CATGAACCCACCCCCCGTACTCCCAGTTCCTGTCCCGGAGGAACGAGATCC 600
    ||||||||||||||| ||||||||||||||||||||||||||||||||||
234 CATGAACCCACCCCCCCTACTCCCAGTTCCTGTCCCGGAGGAACGAGATCC 283
                .         .         .         .         .
601 CCCTAATTCACTTCAACACCCCCATACCACGGCGGCACACCCGGAGCGCC 650
    |||| || |||||||| ||||||| ||||||||||||||||||||||||||
284 CCCTCATCCACTTCAATACCCCCAGACCACGGCGGCACACCCGGAGCGCC 333
                .         .         .         .         .
651 GAGGACGACTCGGAGCGGGACCCCCTGAACGTGCTGAAGCCCCGGGCCCG 700
```

FIGURE 17

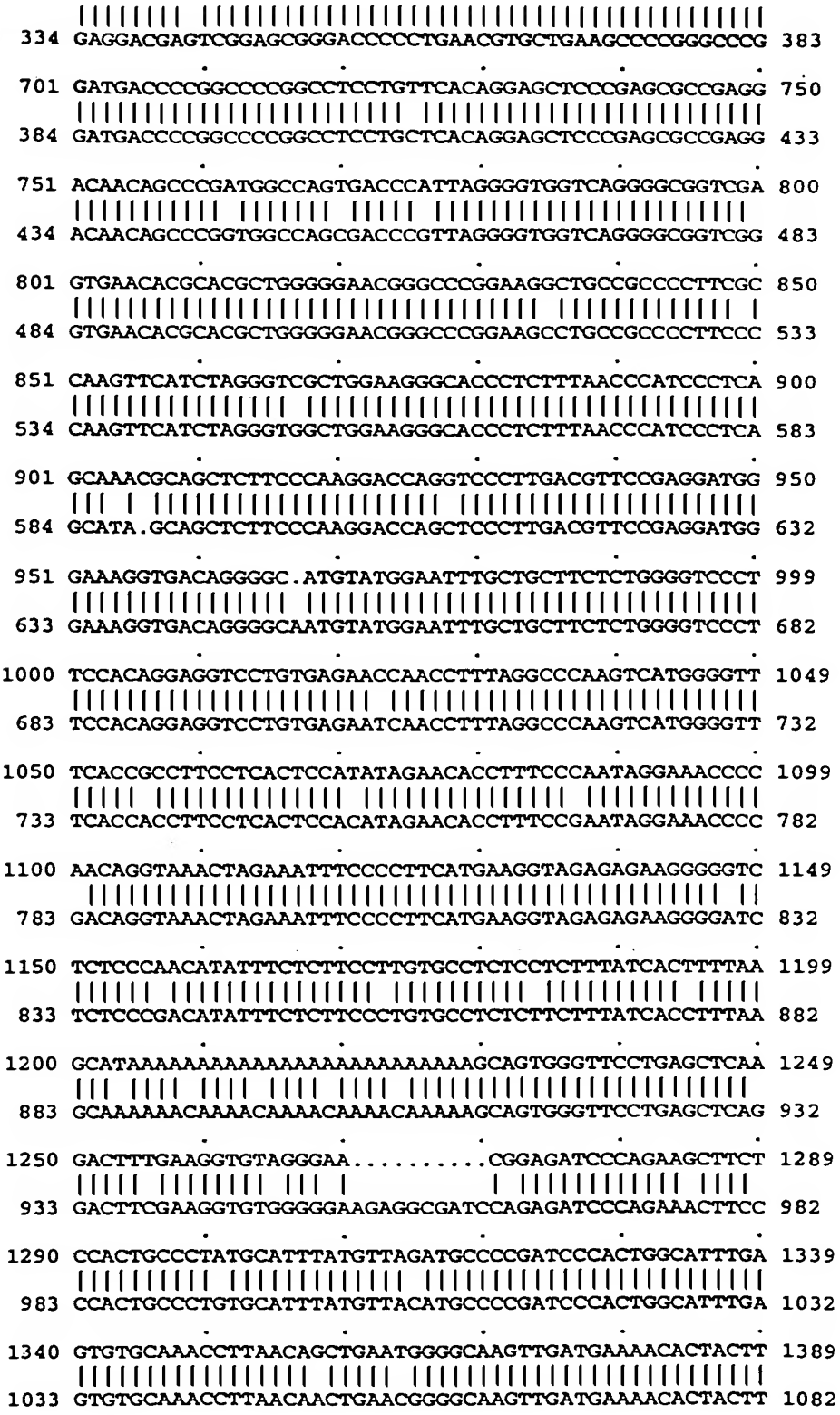

FIGURE 17
CONT.

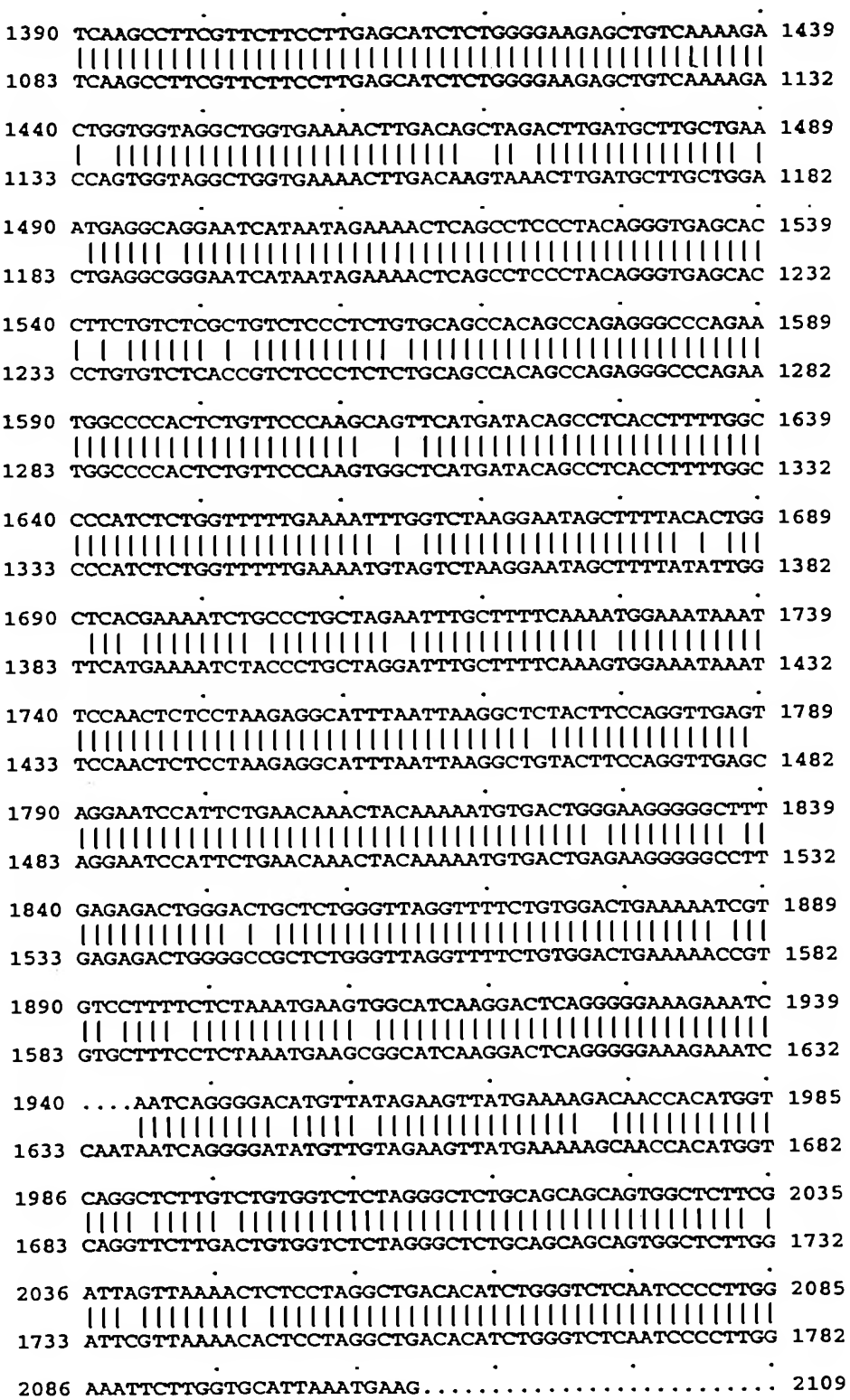

FIGURE 17
CONT.

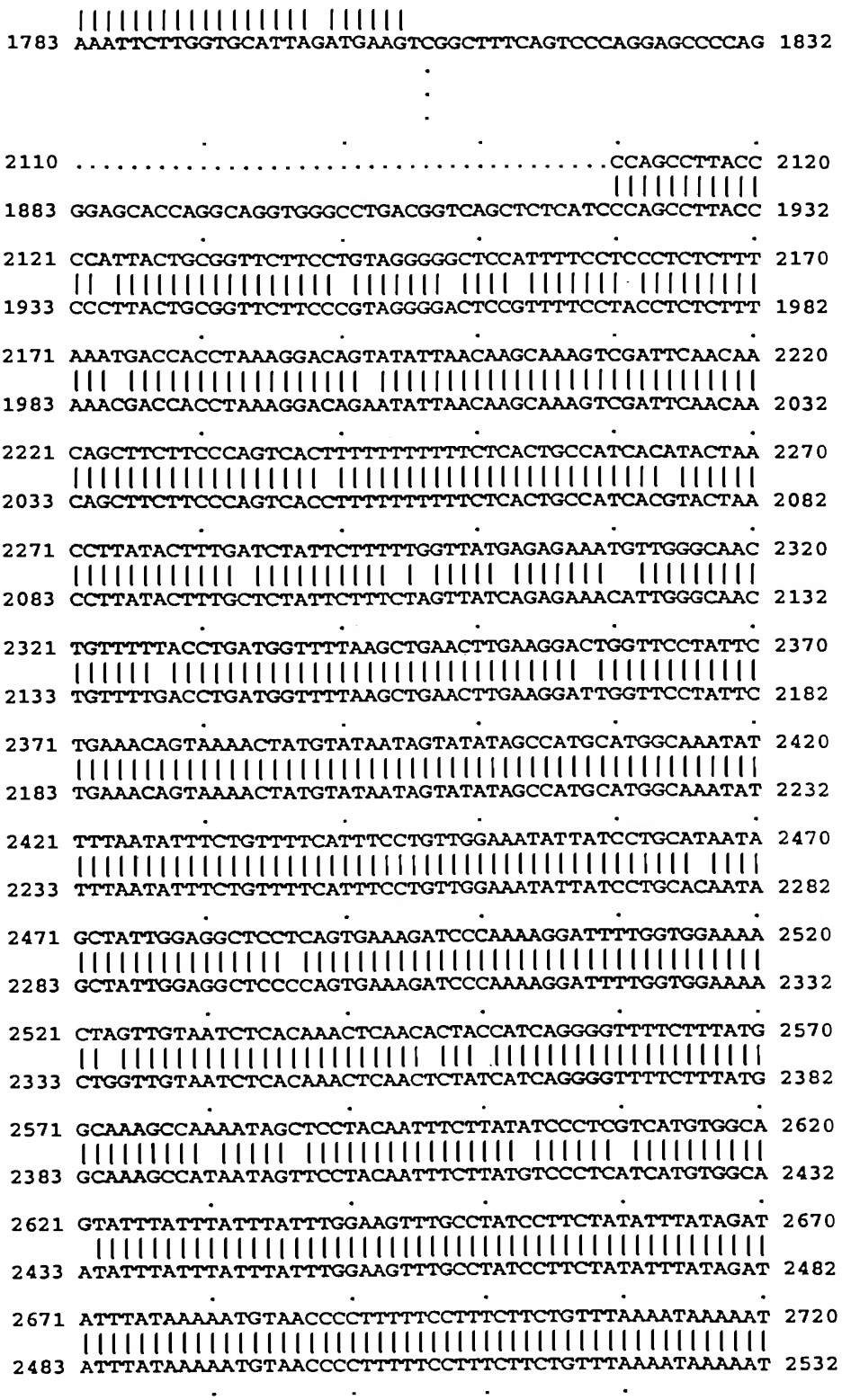

FIGURE 17
CONT.

```
2721 AAAATTTATCTCA.............................. 2733
     |||||||||||||
2533 AAAATTTATCTCAAAAAAAAAAAAAAAAAAAAAAGGGCGGCCGC 2574
```

FIGURE 17
CONT.

```
GAP of: WBAa000fk  check: 851  from: 1  to: 178

human FGF20.pro (analysis only) - Import - complete

to: XBAa000fk  check: 5181  from: 1  to: 181

monkey FGF20.pro (analysis only) - Import - complete

Symbol comparison table: /prod/ddm/seqanal/BLAST/matrix/aa/BLOSUM62
CompCheck: 1102
 Matrix made by matblas from blosum62.iij

       Gap Weight:     12      Average Match:  2.778
    Length Weight:      4   Average Mismatch: -2.248

          Quality:    912             Length:    181
            Ratio:  5.124               Gaps:      0
Percent Similarity: 95.506   Percent Identity: 93.820

      Match display thresholds for the alignment(s):
                  | = IDENTITY
                  : =   2
                  . =   1

WBAa000fk x XBAa000fk

      .         .         .         .
  1 ...MIRSEDAGFVVITGVMSRRYLCMDFRGNIFGSHYFDPENCRFQHQTL 47      hFGF-20
       :||||||||||||||||||||||||||||||||.|||||.| ||
  1 VDPRVRSEDAGFVVITGVMSRRYLCMDFRGNIFGSHYFNPENCRFRHWTL 50      monkeyFGF-20
      .         .         .         .         .
 48 ENGYDVYHSPQYHFLVSLGRAKRAFLPGMNPPPYSQFLSRRNEIPLIHFN 97
    ||||||||||:||||||||||||||||||||||||||||||||||||||||
 51 ENGYDVYHSPQHHFLVSLGRAKRAFLPGMNPPPYSQFLSRRNEIPLIHFN 100
      .         .         .         .         .
 98 TPIPRRHTRSAEDDSERDPLNVLKPRARMTPAPASCSQELPSAEDNSPMA 147
    || ||||||||||:||||||||||||||||||||||||||||||||||.|
101 TPRPRRHTRSAEDESERDPLNVLKPRARMTPAPASCSQELPSAEDNSPVA 150
      .         .         .
148 SDPLGVVRGGRVNTHAGGTGPEGCRPFAKFI 178
    |||||||||||||||||||| |||| |||
151 SDPLGVVRGGRVNTHAGGTGPEACRPFPKFI 181
```

FIGURE 18

SEQUENCE LISTING

<110> Millennium Pharmaceuticals

<120> A Novel Fibroblast Growth Factor Family Member

<130> MNI-081PC

<140>
<141>

<150> 60/127,534
<151> 1999-04-02
<150> 09/454,470
<151> 1999-12-03

<160> 12

<170> PatentIn Ver. 2.0

<210> 1
<211> 805
<212> DNA
<213> Macaca sp.

<220>
<221> CDS
<222> (2)..(532)

<400> 1

```
c gtc cga tca gag gat gct ggc ttt gtg gtg att aca ggt gtg atg agc 49
  Val Arg Ser Glu Asp Ala Gly Phe Val Val Ile Thr Gly Val Met Ser
    1               5                  10                  15

aga aga tac ctc tgc atg gat ttc aga ggc aac att ttt gga tca cac   97
Arg Arg Tyr Leu Cys Met Asp Phe Arg Gly Asn Ile Phe Gly Ser His
            20                  25                  30

tat ttc aac ccg gag aac tgc agg ttc cga cac tgg acg ctg gag aac   145
Tyr Phe Asn Pro Glu Asn Cys Arg Phe Arg His Trp Thr Leu Glu Asn
        35                  40                  45

ggc tac gac gtc tac cac tct cct cag cat cac ttt ctg gtc agt ctg   193
Gly Tyr Asp Val Tyr His Ser Pro Gln His His Phe Leu Val Ser Leu
    50                  55                  60

ggc cgg gcg aag agg gcc ttc ctg cca ggc atg aac cca ccc ccc tac   241
Gly Arg Ala Lys Arg Ala Phe Leu Pro Gly Met Asn Pro Pro Pro Tyr
65                  70                  75                  80

tcc cag ttc ctg tcc cgg agg aac gag atc ccc ctc atc cac ttc aat   289
Ser Gln Phe Leu Ser Arg Arg Asn Glu Ile Pro Leu Ile His Phe Asn
                85                  90                  95

acc ccc aga cca cgg cgg cac acc cgg agc gcc gag gac gag tcg gag   337
Thr Pro Arg Pro Arg Arg His Thr Arg Ser Ala Glu Asp Glu Ser Glu
            100                 105                 110

cgg gac ccc ctg aac gtg ctg aag ccc cgg gcc cgg atg acc ccg gcc   385
```

```
Arg Asp Pro Leu Asn Val Leu Lys Pro Arg Ala Arg Met Thr Pro Ala
        115                 120                 125

ccg gcc tcc tgc tca cag gag ctc ccg agc gcc gag gac aac agc ccg   433
Pro Ala Ser Cys Ser Gln Glu Leu Pro Ser Ala Glu Asp Asn Ser Pro
    130                 135                 140

gtg gcc agc gac ccg tta ggg gtg gtc agg ggc ggt cgg gtg aac acg   481
Val Ala Ser Asp Pro Leu Gly Val Val Arg Gly Gly Arg Val Asn Thr
145                 150                 155                 160

cac gct ggg gga acg ggc ccg gaa gcc tgc cgc ccc ttc ccc aag ttc   529
His Ala Gly Gly Thr Gly Pro Glu Ala Cys Arg Pro Phe Pro Lys Phe
                165                 170                 175

atc tagggtggct ggaagggcac cctctttaac ccatccctca gcatagcaag        582
Ile

ctcttccaag gaccaagctc cttgacgttc cgaggatggg aaaggtgaca ggggcaatgt 642

atggaattgc tgcttctctg gggtcccttc cacaggaggt ccttgtgaga atcaaccttt 702

aggcccaagt catggggttt caacanctttt cttcacttca acatagaaca acctttttccg 762

aataggaaac cccgacaggt aaactagnaa ttttcccctt tat                   805
```

<210> 2
<211> 177
<212> PRT
<213> Macaca sp.

<400> 2

```
Val Arg Ser Glu Asp Ala Gly Phe Val Val Ile Thr Gly Val Met Ser
  1               5                  10                  15

Arg Arg Tyr Leu Cys Met Asp Phe Arg Gly Asn Ile Phe Gly Ser His
             20                  25                  30

Tyr Phe Asn Pro Glu Asn Cys Arg Phe Arg His Trp Thr Leu Glu Asn
         35                  40                  45

Gly Tyr Asp Val Tyr His Ser Pro Gln His His Phe Leu Val Ser Leu
     50                  55                  60

Gly Arg Ala Lys Arg Ala Phe Leu Pro Gly Met Asn Pro Pro Pro Tyr
 65                  70                  75                  80

Ser Gln Phe Leu Ser Arg Arg Asn Glu Ile Pro Leu Ile His Phe Asn
                 85                  90                  95

Thr Pro Arg Pro Arg Arg His Thr Arg Ser Ala Glu Asp Glu Ser Glu
            100                 105                 110

Arg Asp Pro Leu Asn Val Leu Lys Pro Arg Ala Arg Met Thr Pro Ala
        115                 120                 125

Pro Ala Ser Cys Ser Gln Glu Leu Pro Ser Ala Glu Asp Asn Ser Pro
    130                 135                 140
```

```
Val Ala Ser Asp Pro Leu Gly Val Val Arg Gly Gly Arg Val Asn Thr
145                 150                 155                 160

His Ala Gly Gly Thr Gly Pro Glu Ala Cys Arg Pro Phe Pro Lys Phe
                165                 170                 175

Ile
```

<210> 3
<211> 531
<212> DNA
<213> Macaca sp.

<220>
<221> CDS
<222> (1)..(531)

<400> 3

```
gtc cga tca gag gat gct ggc ttt gtg gtg att aca ggt gtg atg agc   48
Val Arg Ser Glu Asp Ala Gly Phe Val Val Ile Thr Gly Val Met Ser
  1               5                  10                  15

aga aga tac ctc tgc atg gat ttc aga ggc aac att ttt gga tca cac   96
Arg Arg Tyr Leu Cys Met Asp Phe Arg Gly Asn Ile Phe Gly Ser His
            20                  25                  30

tat ttc aac ccg gag aac tgc agg ttc cga cac tgg acg ctg gag aac   144
Tyr Phe Asn Pro Glu Asn Cys Arg Phe Arg His Trp Thr Leu Glu Asn
        35                  40                  45

ggc tac gac gtc tac cac tct cct cag cat cac ttt ctg gtc agt ctg   192
Gly Tyr Asp Val Tyr His Ser Pro Gln His His Phe Leu Val Ser Leu
    50                  55                  60

ggc cgg gcg aag agg gcc ttc ctg cca ggc atg aac cca ccc ccc tac   240
Gly Arg Ala Lys Arg Ala Phe Leu Pro Gly Met Asn Pro Pro Pro Tyr
65                  70                  75                  80

tcc cag ttc ctg tcc cgg agg aac gag atc ccc ctc atc cac ttc aat   288
Ser Gln Phe Leu Ser Arg Arg Asn Glu Ile Pro Leu Ile His Phe Asn
                85                  90                  95

acc ccc aga cca cgg cgg cac acc cgg agc gcc gag gac gag tcg gag   336
Thr Pro Arg Pro Arg Arg His Thr Arg Ser Ala Glu Asp Glu Ser Glu
            100                 105                 110

cgg gac ccc ctg aac gtg ctg aag ccc cgg gcc cgg atg acc ccg gcc   384
Arg Asp Pro Leu Asn Val Leu Lys Pro Arg Ala Arg Met Thr Pro Ala
        115                 120                 125

ccg gcc tcc tgc tca cag gag ctc ccg agc gcc gag gac aac agc ccg   432
Pro Ala Ser Cys Ser Gln Glu Leu Pro Ser Ala Glu Asp Asn Ser Pro
    130                 135                 140

gtg gcc agc gac ccg tta ggg gtg gtc agg ggc ggt cgg gtg aac acg   480
Val Ala Ser Asp Pro Leu Gly Val Val Arg Gly Gly Arg Val Asn Thr
```

```
145                 150                 155                 160

cac gct ggg gga acg ggc ccg gaa gcc tgc cgc ccc ttc ccc aag ttc   528
His Ala Gly Gly Thr Gly Pro Glu Ala Cys Arg Pro Phe Pro Lys Phe
                165                 170                 175

atc                                                               531
Ile
```

<210> 4
<211> 2749
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (326)..(862)

<400> 4

```
gggaaaataa agacggtcca ggtagagaga gagaaacatg tgttcagcac aggtagaaga 60

attccaggag ctcagagtgc cccatacagg caacaagatg aagcaggagg tgaatgactg 120

tatgtgtgtt gggggcaaga gaggatgtca gaagaaacgc tgaatatgca gaaatgaggc 180

tgaatttaag agtgctgaag ttatcaccac ccttaaaatc aatccaggga ggtttcatga 240

aggtaggttt tcaggaggtg cttgaaggtg ggaattggat ggcaatgagt ctttgccctg 300

cctgtttttc tccataggtg ccctg atg atc aga tca gag gat gct ggc ttt   352
                          Met Ile Arg Ser Glu Asp Ala Gly Phe
                            1               5

gtg gtg att aca ggt gtg atg agc aga aga tac ctc tgc atg gat ttc   400
Val Val Ile Thr Gly Val Met Ser Arg Arg Tyr Leu Cys Met Asp Phe
 10                  15                  20                  25

aga ggc aac att ttt gga tca cac tat ttc gac ccg gag aac tgc agg   448
Arg Gly Asn Ile Phe Gly Ser His Tyr Phe Asp Pro Glu Asn Cys Arg
                30                  35                  40

ttc caa cac cag acg ctg gaa aac ggg tac gac gtc tac cac tct cct   496
Phe Gln His Gln Thr Leu Glu Asn Gly Tyr Asp Val Tyr His Ser Pro
            45                  50                  55
cag tat cac ttc ctg gtc agt ctg ggc cgg gcg aag aga gcc ttc ctg   544
Gln Tyr His Phe Leu Val Ser Leu Gly Arg Ala Lys Arg Ala Phe Leu
        60                  65                  70

cca ggc atg aac cca ccc ccg tac tcc cag ttc ctg tcc cgg agg aac   592
Pro Gly Met Asn Pro Pro Pro Tyr Ser Gln Phe Leu Ser Arg Arg Asn
    75                  80                  85

gag atc ccc cta att cac ttc aac acc ccc ata cca cgg cgg cac acc   640
Glu Ile Pro Leu Ile His Phe Asn Thr Pro Ile Pro Arg Arg His Thr
 90                  95                  100                 105

cgg agc gcc gag gac gac tcg gag cgg gac ccc ctg aac gtg ctg aag   688
```

```
Arg Ser Ala Glu Asp Asp Ser Glu Arg Asp Pro Leu Asn Val Leu Lys
                110                 115                 120

ccc cgg gcc cgg atg acc ccg gcc ccg gcc tcc tgt tca cag gag ctc   736
Pro Arg Ala Arg Met Thr Pro Ala Pro Ala Ser Cys Ser Gln Glu Leu
            125                 130                 135

ccg agc gcc gag gac aac agc ccg atg gcc agt gac cca tta ggg gtg   784
Pro Ser Ala Glu Asp Asn Ser Pro Met Ala Ser Asp Pro Leu Gly Val
        140                 145                 150

gtc agg ggc ggt cga gtg aac acg cac gct ggg gga acg ggc ccg gaa   832
Val Arg Gly Gly Arg Val Asn Thr His Ala Gly Gly Thr Gly Pro Glu
    155                 160                 165

ggc tgc cgc ccc ttc gcc aag ttc atc tag ggtcgctgga agggcaccct     882
Gly Cys Arg Pro Phe Ala Lys Phe Ile
170                 175

ctttaaccca tccctcagca aacgcagctc ttcccaagga ccaggtccct tgacgttccg 942

aggatgggaa aggtgacagg ggcatgtatg gaatttgctg cttctctggg gtcccttcca 1002

caggaggtcc tgtgagaacc aacctttgag gcccaagtca tggggtttca ccgccttcct 1062

cactccatat agaacacctt tcccaatagg aaaccccaac aggtaaacta gaaatttccc 1122

cttcatgaag gtagagagaa ggggtctctc ccaacatatt tctcttcctt gtgcctctcc 1182

tctttatcac ttttaagcat aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aagcagtggg 1242

ttcctgagct caagactttg aaggtgtagg gaagaggaaa tcggagatcc cagaagcttc 1302

tccactgccc tatgcattta tgttagatgc cccgatccca ctggcatttg agtgtgcaaa 1362

ccttgacatt aacagctgaa tggggcaagt tgatgaaaac actactttca agccttcgtt 1422

cttccttgag catctctggg gaagagctgt caaaagactg gtggtaggct ggtgaaaact 1482

tgacagctag acttgatgct tgctgaaatg aggcaggaat cataatagaa aactcagcct 1542

ccctacaggg tgagcacctt ctgtctcgct gtctccctct gtgcagccac agccagaggg 1602

cccagaatgg ccccactctg ttcccaagca gttcatgata cagcctcacc ttttggcccc 1662

atctctggtt tttgaaaatt tggtctaagg aataaatagc ttttacactg gctcacgaaa 1722

atctgccctg ctagaatttg cttttcaaaa tggaaataaa ttccaactct cctaagaggc 1782

atttaattaa ggctctactt ccaggttgag taggaatcca ttctgaacaa actacaaaaa 1842

tgtgactggg aaggggggctt tgagagactg ggactgctct gggttaggtt ttctgtggac 1902

tgaaaaatcg tgtccttttc tctaaatgaa gtggcatcaa ggactcaggg ggaaagaaat 1962

caggggacat gttatagaag ttatgaaaag acaaccacat ggtcaggctc ttgtctgtgg 2022

tctctagggc tctgcagcag cagtggctct tcgattagtt aaaactctcc taggctgaca 2082
```

catctgggtc tcaatcccct tggaaattct tggtgcatta aatgaagcct taccccatta 2142

ctgcggttct tcctgtaagg gggctccatt ttcctccctc tctttaaatg accacctaaa 2202

ggacagtata ttaacaagca aagtcgattc aacaacagct tcttcccagt cacttttttt 2262

tttctcactg ccatcacata ctaaccttat actttgatct attcttttg gttatgagag 2322

aaatgttggg caactgtttt tacctgatgg ttttaagctg aacttgaagg actggttcct 2382

attctgaaac agtaaaacta tgtataatag tatatagcca tgcatggcaa atattttaat 2442

atttctgttt tcatttcctg ttggaaatat tatcctgcat aatagctatt ggaggctcct 2502

cagtgaaaga tcccaaaagg attttggtgg aaaactagtt gtaatctcac aaactcaaca 2562

ctaccatcag gggttttctt tatggcaaag ccaaaatagc tcctacaatt tcttatatcc 2622

ctcgtcatgt ggcagtattt atttatttat ttggaagttt gcctatcctt ctatatttat 2682

agatatttat aaaaatgtaa ccccttttc ctttcttctg tttaaaataa aaataaaatt 2742

tatctca 2749

<210> 5
<211> 178
<212> PRT
<213> Homo sapiens

<400> 5

```
Met Ile Arg Ser Glu Asp Ala Gly Phe Val Val Ile Thr Gly Val Met
  1               5                  10                  15

Ser Arg Arg Tyr Leu Cys Met Asp Phe Arg Gly Asn Ile Phe Gly Ser
             20                  25                  30

His Tyr Phe Asp Pro Glu Asn Cys Arg Phe Gln His Gln Thr Leu Glu
         35                  40                  45

Asn Gly Tyr Asp Val Tyr His Ser Pro Gln Tyr His Phe Leu Val Ser
     50                  55                  60

Leu Gly Arg Ala Lys Arg Ala Phe Leu Pro Gly Met Asn Pro Pro Pro
 65                  70                  75                  80

Tyr Ser Gln Phe Leu Ser Arg Arg Asn Glu Ile Pro Leu Ile His Phe
                 85                  90                  95

Asn Thr Pro Ile Pro Arg Arg His Thr Arg Ser Ala Glu Asp Asp Ser
            100                 105                 110

Glu Arg Asp Pro Leu Asn Val Leu Lys Pro Arg Ala Arg Met Thr Pro
        115                 120                 125

Ala Pro Ala Ser Cys Ser Gln Glu Leu Pro Ser Ala Glu Asp Asn Ser
    130                 135                 140

Pro Met Ala Ser Asp Pro Leu Gly Val Val Arg Gly Gly Arg Val Asn
```

145 150 155 160

Thr His Ala Gly Gly Thr Gly Pro Glu Gly Cys Arg Pro Phe Ala Lys
165 170 175

Phe Ile

<210> 6
<211> 537
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(537)

<400> 6

```
atg atc aga tca gag gat gct ggc ttt gtg gtg att aca ggt gtg atg   48
Met Ile Arg Ser Glu Asp Ala Gly Phe Val Val Ile Thr Gly Val Met
  1               5                  10                  15

agc aga aga tac ctc tgc atg gat ttc aga ggc aac att ttt gga tca   96
Ser Arg Arg Tyr Leu Cys Met Asp Phe Arg Gly Asn Ile Phe Gly Ser
             20                  25                  30

cac tat ttc gac ccg gag aac tgc agg ttc caa cac cag acg ctg gaa   144
His Tyr Phe Asp Pro Glu Asn Cys Arg Phe Gln His Gln Thr Leu Glu
         35                  40                  45

aac ggg tac gac gtc tac cac tct cct cag tat cac ttc ctg gtc agt   192
Asn Gly Tyr Asp Val Tyr His Ser Pro Gln Tyr His Phe Leu Val Ser
     50                  55                  60

ctg ggc cgg gcg aag aga gcc ttc ctg cca ggc atg aac cca ccc ccg   240
Leu Gly Arg Ala Lys Arg Ala Phe Leu Pro Gly Met Asn Pro Pro Pro
 65                  70                  75                  80

tac tcc cag ttc ctg tcc cgg agg aac gag atc ccc cta att cac ttc   288
Tyr Ser Gln Phe Leu Ser Arg Arg Asn Glu Ile Pro Leu Ile His Phe
                 85                  90                  95

aac acc ccc ata cca cgg cgg cac acc cgg agc gcc gag gac gac tcg   336
Asn Thr Pro Ile Pro Arg Arg His Thr Arg Ser Ala Glu Asp Asp Ser
            100                 105                 110

gag cgg gac ccc ctg aac gtg ctg aag ccc cgg gcc cgg atg acc ccg   384
Glu Arg Asp Pro Leu Asn Val Leu Lys Pro Arg Ala Arg Met Thr Pro
        115                 120                 125

gcc ccg gcc tcc tgt tca cag gag ctc ccg agc gcc gag gac aac agc   432
Ala Pro Ala Ser Cys Ser Gln Glu Leu Pro Ser Ala Glu Asp Asn Ser
    130                 135                 140

ccg atg gcc agt gac cca tta ggg gtg gtc agg ggc ggt cga gtg aac   480
Pro Met Ala Ser Asp Pro Leu Gly Val Val Arg Gly Gly Arg Val Asn
145                 150                 155                 160

acg cac gct ggg gga acg ggc ccg gaa ggc tgc cgc ccc ttc gcc aag   528
```

Thr His Ala Gly Gly Thr Gly Pro Glu Gly Cys Arg Pro Phe Ala Lys
165 170 175

ttc atc tag 537
Phe Ile

<210> 7
<211> 1973
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1071)..(1604)

<400> 7
ccacgcgtcc ggtggggaag aaatctcgct gaattatcac gcatgttaca ccagtatatg 60

atctaattgt gcctttgcca caaaacagta atttaaagcc attatcaatt acttaagagg 120

taggtcgtgt gaatgggttt caggcccttg tcggagacta gtttttgaga ggggacactg 180

aaagtccatg aggggctgca cctggagagg tcaccaccaa gtgagaaaat gacaaagaac 240

caacccaaga agagccaaga agaaaattcc atccgtcact tatattgatt caacataaac 300

agttataccc tctgctccta agcagctcac tctaaggaac gcactggata ggtaaactca 360

gctaaagcaa gttaaatgga atacatgctg taatagaggt gaaggcattg tcctgaggag 420

ctgagaagga agaacaactg attttgaatg gaaagatgag gaaagtcttc atagagatgg 480

tgacgcctga gcctggtctt gaagagtgag tgacttcaat aagtagagaa ggaagaggga 540

gatcaactct actaccattc tgtacacata ctgggtgttg actgatgtat tagacaatta 600

cacagacatc caggaggaga atcagactct atggcaagct ggatccttga aagacatctc 660

agcatagatt taaaaatcac aaagtagaag gcatggaaga atgtgactat caccacaaac 720

attcaaaggt attagtaagg caaaagggaa aataaagacg gtccaggtag agagagagaa 780

acatgtgttc agcacaggta gaagaattcc aggagctcag agtgccccat acaggcaaca 840

agatgaagca ggaggtgaat gactgtatgt gtgttggggg caagagagga tgtcagaaga 900

aacgctgaat atgcagaaat gaggctgaat ttaagagtgc tgaagttatc accaccctta 960

aaatcaatcc agggaggttt catgaaggta ggttttcagg aggtgcttga aggtgggaat 1020

tggatggcaa tgagtctttg ccctgcctgt ttttctccat aggtgccctg atg atc 1076
Met Ile
1

aga tca gag gat gct ggc ttt gtg gtg att aca ggt gtg atg agc aga 1124
Arg Ser Glu Asp Ala Gly Phe Val Val Ile Thr Gly Val Met Ser Arg
5 10 15

```
aga tac ctc tgc atg gat ttc aga ggc aac att ttt gga tca cac tat   1172
Arg Tyr Leu Cys Met Asp Phe Arg Gly Asn Ile Phe Gly Ser His Tyr
     20                  25                  30

ttc gac ccg gag aac tgc agg ttc caa cac cag acg ctg gaa aac ggg   1220
Phe Asp Pro Glu Asn Cys Arg Phe Gln His Gln Thr Leu Glu Asn Gly
 35                  40                  45                  50

tac gac gtc tac cac tct cct cag tat cac ttc ctg gtc agt ctg ggc   1268
Tyr Asp Val Tyr His Ser Pro Gln Tyr His Phe Leu Val Ser Leu Gly
                 55                  60                  65

cgg gcg aag aga gcc ttc ctg cca ggc atg aac cca ccc ccg tac tcc   1316
Arg Ala Lys Arg Ala Phe Leu Pro Gly Met Asn Pro Pro Pro Tyr Ser
             70                  75                  80

cag ttc ctg tcc cgg agg aac gag atc ccc cta att cac ttc aac acc   1364
Gln Phe Leu Ser Arg Arg Asn Glu Ile Pro Leu Ile His Phe Asn Thr
         85                  90                  95

ccc ata cca cgg cgg cac acc cgg agc gcc gag gac gac tcg gag cgg   1412
Pro Ile Pro Arg Arg His Thr Arg Ser Ala Glu Asp Asp Ser Glu Arg
    100                 105                 110

gac ccc ctg aac gtg ctg aag ccc cgg gcc cgg atg acc ccg gcc ccg   1460
Asp Pro Leu Asn Val Leu Lys Pro Arg Ala Arg Met Thr Pro Ala Pro
115                 120                 125                 130

gcc tcc tgt tca cag gag ctc ccg agc gcc gag gac aac agc ccg atg   1508
Ala Ser Cys Ser Gln Glu Leu Pro Ser Ala Glu Asp Asn Ser Pro Met
                135                 140                 145

gcc agt gac cca tta ggg gtg gtc agg ggc ggt cga gtg aac acg cac   1556
Ala Ser Asp Pro Leu Gly Val Val Arg Gly Gly Arg Val Asn Thr His
            150                 155                 160

gct ggg gga acg ggc ccg gaa ggc tgc cgc ccc ttc gcc aag ttc atc   1604
Ala Gly Gly Thr Gly Pro Glu Gly Cys Arg Pro Phe Ala Lys Phe Ile
        165                 170                 175

tagggtcgct ggaagggcac cctctttaac ccatccctca gcaaacgcag ctcttcccaa 1664

ggaccaggtc ccttgacgtt ccgaggatgg gaaaggtgac aggggcatgt atggaatttg 1724

ctgcttctct ggggtccctt ccacaggagg tcctgtgaga accaaccttt gaggcccaag 1784

tcatggggtt tcaccgcctt cctcactcca tatagaacac ctttcccaat aggaaacccc 1844

aacaggtaaa ctagaaattt ccccttcatg aaggtagaga gaaggggtct ctcccaacat 1904

atttctcttc cttgtgcctc tcctctttat cacttttaag cataaaaaaa aaaaaaaaaa 1964

aaaaaaaaa                                                         1973
```

<210> 8
<211> 178
<212> PRT
<213> Homo sapiens

<400> 8
Met Ile Arg Ser Glu Asp Ala Gly Phe Val Val Ile Thr Gly Val Met
1 5 10 15

Ser Arg Arg Tyr Leu Cys Met Asp Phe Arg Gly Asn Ile Phe Gly Ser
20 25 30

His Tyr Phe Asp Pro Glu Asn Cys Arg Phe Gln His Gln Thr Leu Glu
35 40 45

Asn Gly Tyr Asp Val Tyr His Ser Pro Gln Tyr His Phe Leu Val Ser
50 55 60

Leu Gly Arg Ala Lys Arg Ala Phe Leu Pro Gly Met Asn Pro Pro Pro
65 70 75 80

Tyr Ser Gln Phe Leu Ser Arg Arg Asn Glu Ile Pro Leu Ile His Phe
85 90 95

Asn Thr Pro Ile Pro Arg Arg His Thr Arg Ser Ala Glu Asp Asp Ser
100 105 110

Glu Arg Asp Pro Leu Asn Val Leu Lys Pro Arg Ala Arg Met Thr Pro
115 120 125

Ala Pro Ala Ser Cys Ser Gln Glu Leu Pro Ser Ala Glu Asp Asn Ser
130 135 140

Pro Met Ala Ser Asp Pro Leu Gly Val Val Arg Gly Gly Arg Val Asn
145 150 155 160

Thr His Ala Gly Gly Thr Gly Pro Glu Gly Cys Arg Pro Phe Ala Lys
165 170 175

Phe Ile

<210> 9
<211> 534
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(534)

<400> 9
atg atc aga tca gag gat gct ggc ttt gtg gtg att aca ggt gtg atg 48
Met Ile Arg Ser Glu Asp Ala Gly Phe Val Val Ile Thr Gly Val Met
1 5 10 15

agc aga aga tac ctc tgc atg gat ttc aga ggc aac att ttt gga tca 96
Ser Arg Arg Tyr Leu Cys Met Asp Phe Arg Gly Asn Ile Phe Gly Ser
20 25 30

cac tat ttc gac ccg gag aac tgc agg ttc caa cac cag acg ctg gaa 144
His Tyr Phe Asp Pro Glu Asn Cys Arg Phe Gln His Gln Thr Leu Glu

```
         35                  40                  45

aac ggg tac gac gtc tac cac tct cct cag tat cac ttc ctg gtc agt   192
Asn Gly Tyr Asp Val Tyr His Ser Pro Gln Tyr His Phe Leu Val Ser
     50                  55                  60

ctg ggc cgg gcg aag aga gcc ttc ctg cca ggc atg aac cca ccc ccg   240
Leu Gly Arg Ala Lys Arg Ala Phe Leu Pro Gly Met Asn Pro Pro Pro
 65                  70                  75                  80

tac tcc cag ttc ctg tcc cgg agg aac gag atc ccc cta att cac ttc   288
Tyr Ser Gln Phe Leu Ser Arg Arg Asn Glu Ile Pro Leu Ile His Phe
                85                  90                  95

aac acc ccc ata cca cgg cgg cac acc cgg agc gcc gag gac gac tcg   336
Asn Thr Pro Ile Pro Arg Arg His Thr Arg Ser Ala Glu Asp Asp Ser
            100                 105                 110

gag cgg gac ccc ctg aac gtg ctg aag ccc cgg gcc cgg atg acc ccg   384
Glu Arg Asp Pro Leu Asn Val Leu Lys Pro Arg Ala Arg Met Thr Pro
        115                 120                 125

gcc ccg gcc tcc tgt tca cag gag ctc ccg agc gcc gag gac aac agc   432
Ala Pro Ala Ser Cys Ser Gln Glu Leu Pro Ser Ala Glu Asp Asn Ser
    130                 135                 140

ccg atg gcc agt gac cca tta ggg gtg gtc agg ggc ggt cga gtg aac   480
Pro Met Ala Ser Asp Pro Leu Gly Val Val Arg Gly Gly Arg Val Asn
145                 150                 155                 160

acg cac gct ggg gga acg ggc ccg gaa ggc tgc cgc ccc ttc gcc aag   528
Thr His Ala Gly Gly Thr Gly Pro Glu Gly Cys Arg Pro Phe Ala Lys
                165                 170                 175

ttc atc                                                          534
Phe Ile
```

<210> 10
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
oligonucleotide primer

<400> 10
gccttcctgc caggcatgaa cc 22

<210> 11
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
oligonucleotide primer

<400> 11
ctttcccatc ctcggaacgt caagg 25

<210> 12
<211> 25
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: consensus pattern

<220>
<223> Xaa's at postions 2,4,6-12,15,17, and 19-24 may be any amino acid

<220>
<223> Xaa at position 3 may be Lys or Ile

<220>
<223> Xaa at position 5 may be Ser, Thr, Ala, Gly or Pro

<220>
<223> Any one of the Xaa's between postions 6-12 may be absent; intended to equal a range of 6-7 amino acids

<220>
<223> Xaa at position 13 may be Asp or Glu

<220>
<223> Xaa at postion 16 may be Phe, Leu or Met

<400> 12
Gly Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa
1 5 10 15

Xaa Glu Xaa Xaa Xaa Xaa Xaa Xaa Tyr
20 25

INTERNATIONAL SEARCH REPORT

Interr nal Application No
PCT/US 00/08076

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/16 C07K14/50 C07K16/22 A61K38/22 G01N33/53
G01N33/68 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
BIOSIS, EPO-Internal, SCISEARCH, EMBASE, BIOTECHNOLOGY ABS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ZHAO S. ET AL.: "Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building" EMBL DATABASE ; ACCESSION NUMBER AQ412342, 25 March 1999 (1999-03-25), XP002141756 the whole document ---	5-11
A	NISHIMURA ET AL: "Structure and expression of a novel human FGF, FGF-19, expressed in fetal brain" BIOCHIMICA ET BIOPHYSICA ACTA,NL,AMSTERDAM, vol. 1444, no. 1444, 18 January 1999 (1999-01-18), pages 148-151-51, XP002099435 ISSN: 0006-3002 cited in the application the whole document --- -/--	1,9,11, 12,16, 17,19

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
4 July 2000	14/07/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Gurdjian, D

1

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inter nal Application No
PCT/US 00/08076

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	WO 96 39509 A (HUMAN GENOME SCIENCES INC ;GREENE JOHN M (US); ROSEN CRAIG A (US)) 12 December 1996 (1996-12-12) the whole document	1,9,11, 12,16, 17,19

1

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 3

As no accesion number is given , the search was not carried out

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

...ormation on patent family members

Interr... ...al Application No
PCT/US 00/08076

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9639509 A	12-12-1996	AU 711647 B	21-10-1999
		AU 2767495 A	24-12-1996
		EP 0832216 A	01-04-1998
		JP 11506919 T	22-06-1999